

Broadcast Gossip Algorithms for Consensus on Strongly Connected Digraphs

Wu Shaochuan and Michael G. Rabbat

September 3, 2012

Abstract

We study a general framework for broadcast gossip algorithms which use companion variables to solve the average consensus problem. Each node maintains an initial state and a companion variable. Iterative updates are performed asynchronously whereby one random node broadcasts its current state and companion variable and all other nodes receiving the broadcast update their state and companion variable. We provide conditions under which this scheme is guaranteed to converge to a consensus solution, where all nodes have the same limiting values, on any strongly connected directed graph. Under stronger conditions, which are reasonable when the underlying communication graph is undirected, we guarantee that the consensus value is equal to the average, both in expectation and in the mean-squared sense. Our analysis uses tools from non-negative matrix theory and perturbation theory. The perturbation results rely on a parameter being sufficiently small. We characterize the allowable upper bound as well as the optimal setting for the perturbation parameter as a function of the network topology, and this allows us to characterize the worst-case rate of convergence. Simulations illustrate that, in comparison to existing broadcast gossip algorithms, the approaches proposed in this paper have the advantage that they simultaneously can be guaranteed to converge to the average consensus and they converge in a small number of broadcasts.

1 Introduction

Gossip algorithms are an attractive solution for information processing in applications such as distributed signal processing [1], networked control [2], and multi-robot systems [3]. They are attractive because they require little infrastructure; nodes iteratively pass messages with their immediate neighbors in a network until they reach a consensus on the solution. Consequently, there is little overhead associated with forming and maintaining specialized routes, and there are no bottlenecks or single points of failure.

Broadcast gossip algorithms, introduced in [4–6], are especially attractive for use in wireless networks. Unlike the majority of existing gossip algorithms, where messages are either asynchronously exchanged between pairs of nodes or where nodes synchronously exchange and process messages

Wu Shaochuan is with the Department of Electrical and Information Engineering, Harbin Institute of Technology, Harbin, 150080 China e-mail: scwu@hit.edu.cn. This work was conducted while W.S. was visiting McGill University.

Michael G. Rabbat is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 0E9 Canada, e-mail: michael.rabbat@mcgill.ca.

The work of W.S. was partially supported by the National Science Foundation of China (NSFC) grant No. 61201147. The work of M.G.R. was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the *Ministère du Développement économique, de l'Innovation et de l'Exportation* (MDEIE).

with all of their neighbors, in broadcast gossip algorithms nodes asynchronously broadcast a message and the message contents are immediately processed by all neighbors receiving it. By exploiting the broadcast nature of wireless communications, broadcast gossip algorithms are more efficient (they converge after fewer transmissions) than other gossip algorithms [6]. However, previously proposed broadcast gossip algorithms either converge to a consensus on a random solution [6], which may not be acceptable in practical applications, or they do not have theoretical guarantees [7].

In this article we propose and analyze a family of broadcast gossip algorithms for strongly connected directed graphs. If the network is symmetric (undirected) or if nodes know their out-degree, these algorithms are guaranteed to converge to the average consensus both in expectation and in the mean-squared sense. In more general settings, the algorithms are still guaranteed to converge to a specific solution which is a convex combination of the initial values at all nodes in the network (but not necessarily the average). We give a precise characterization of this solution in terms of the algorithm parameters. Our analysis combines tools and techniques from non-negative matrix theory and matrix perturbation theory. Along these lines, we derive an upper bound on the perturbation parameter under which convergence is guaranteed, and we derive an expression for the optimal value of the perturbation parameter.

1.1 Related Work

Broadcast gossip algorithms (BGAs) are introduced in the series of papers by Aysal et al. [4–6]. The BGAs proposed there involve nodes asynchronously transmitting a scalar-valued message, and each time a node receives a message from its neighbors it performs an update by forming a convex combination of the received value with their own previous value. Then, when it is a given node’s turn to broadcast next (as determined by a random timer, in the asynchronous model [8, 9]), the node broadcasts its current value. In [4–6] it is shown that, when executed over an undirected graph (i.e., one with symmetric links) such an algorithm converges to a consensus solution almost surely. The updates of this algorithm are linear and can be expressed as a random, time-varying matrix acting on the vector containing the state values at each node. Unlike conventional pairwise or synchronous gossip algorithms, the matrices in [4–6] corresponding to the update when a particular node transmits cannot be viewed as the transition matrix of a reversible Markov chain, and so the average is not preserved from iteration to iteration. Consequently, although the consensus value is equal to the average of the initial values at every node in expectation, for any particular sample path (where the randomness is in the sequence determining the order in which nodes broadcast) the consensus value is randomly distributed about the average of their initial values but is not precisely equal to it.

Subsequent recent work [10] investigates related BGAs, demonstrating that their convergence properties are robust even when the broadcasts from different nodes may interfere at a receiver. A broadcast-based algorithm has also been proposed for solving distributed convex optimization problems [11].

A modified BGA is proposed by Franceschelli et al. [7, 12], where nodes maintain a companion (or surplus) variable in addition to the state variable they seek to average. By careful accounting of both the companion and state variables, a conservation principle is established, and simulation results suggest that the algorithm with companion variables converges to the average consensus for all sample paths, not just in expectation. However, no proof of convergence or theoretical convergence rate analysis is available for the algorithm of [7, 12].

Recent work of Cai and Ishii [13, 14] analyzes related distributed averaging algorithms on directed graphs that use companion variables. The two types of algorithms analyzed in [13, 14] involve asynchronous pairwise updates and synchronous updates. They make use of tools from

matrix perturbation theory, and the work in the present article can be seen as generalizing the results in [13, 14] for broadcast gossip updates.

1.2 Contributions and Paper Organization

The contributions of this article are as follows. In Section 2 we propose a general framework for broadcast gossip algorithms over directed graphs using companion variables. For this framework we determine conditions on the algorithm parameters under which convergence to a consensus is guaranteed both in expectation (in Section 3) and in the mean squared sense (in Section 4).

We then consider two specific instances of the general framework in Sections 5 and 6. In one instance, which we refer to as unbiased broadcast gossip algorithms (cf. Section 5), the consensus value is guaranteed to be the average of the initial values. In the other instance (biased broadcast gossip algorithms, Sec. 6), the consensus value is no longer the average of the initial values, but it depends on the stationary distribution of a Markov chain associated with the algorithm parameters. The unbiased algorithm requires that each node be aware of its out-degree, the number of nodes that receive its broadcasts. This is a reasonable assumption in networks where connectivity is symmetric, but it may not be reasonable in networks with directed edges. In particular, if there are directed edges, then there is no immediate feedback link, making it more challenging for a node to identify the out-neighbors that receive its broadcasts. This motivates further study of the biased BGAs, which are more practical in such scenarios because they do not require that nodes know their out-degree.

Our analysis of the general framework makes use of tools from matrix perturbation theory. In particular, the way in which the information in the companion variables is incorporated back into the main state variables depends on a parameter which can be viewed as controlling the extent to which a baseline linear system is perturbed. For sufficiently small values of the perturbation parameter, the algorithm is guaranteed to converge. In Section 7 we determine a tight upper bound on the allowable values for the perturbation parameter for biased broadcast gossip algorithms. This bound constitutes an improvement over previous bounds along these lines because it explicitly takes into account the structure of the graph through spectral properties of a corresponding graph Laplacian matrix. In addition to determining this bound, we identify a topology-dependent optimal value for the perturbation parameter in Section 8, and we obtain an expression for the resulting second largest eigenvalue which governs the worst-case rate of convergence.

Simulation results, reported in Section 9, demonstrate that the proposed broadcast gossip algorithms fare well compared to the existing algorithms [6, 7]. The algorithm of [6] converges quickly but can converge to a consensus value which is very far from the average. The algorithm of [7] converges to the average consensus but requires significantly more iterations than the algorithm of [6]. The algorithms proposed here converge quickly and they can be made to converge to the average consensus. We conclude in Section 10.

1.3 Notation

Before proceeding, we summarize some of the notation used in this article. Let $x \in \mathbb{R}^n$ be a n -dimensional column vector. The Euclidean norm of x is denoted by $\|x\|_2$. Let A be a $n \times n$ matrix with real-valued entries. Let $[A]_{i,j}$ denote the entry in the i th row and j th column of A ; we also write $A_{i,j}$ when there is no ambiguity. The ∞ -norm of A is given by $\|A\|_\infty = \max_i \sum_{j=1}^n |A_{i,j}|$, the largest absolute row sum, and the 1-norm of A is given by $\|A\|_1 = \max_j \sum_{i=1}^n |A_{i,j}|$, the largest absolute column sum. The spectral radius of A is the largest modulus of an eigenvalue of A and is denoted by $\rho(A) = \max_{i=1,\dots,n} |\lambda_i(A)|$, where $\lambda_1(A), \dots, \lambda_n(A)$ are the eigenvalues of A . For the

vector $x \in \mathbb{R}^n$, let $\text{diag}(x)$ denote a $n \times n$ diagonal matrix with $[\text{diag}(x)]_{i,i} = x_i$. For a $n \times n$ matrix A , let $\text{diag}(A)$ denote a n -dimensional column vector with $[\text{diag}(A)]_i = A_{i,i}$.

2 Framework for Broadcast Gossip Algorithms

2.1 Network Model

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph which represents the network connectivity, where $\mathcal{V} = \{1, \dots, n\}$ is the set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges. The network contains a directed edge $(i, j) \in \mathcal{E}$ if and only if node i receives messages transmitted by node j . Let $\mathcal{N}_i^+ = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$ and $\mathcal{N}_i^- = \{k \in \mathcal{V} : (k, i) \in \mathcal{E}\}$ denote the set of in-neighbors and out-neighbors, respectively, of node i . For the rest of this paper we make the following assumption.

Assumption 1. *The graph \mathcal{G} is strongly connected; i.e., for any pair of nodes $i, j \in \mathcal{V}$, there exists a sequence of nodes $i = i_0, i_1, i_2, \dots, i_m = j$ such that $(i_{\ell-1}, i_\ell) \in \mathcal{E}$ for all $\ell = 1, \dots, m$.*

2.2 Distributed Averaging

The goal of any broadcast gossip algorithm is to accomplish distributed averaging. Each node $i \in \mathcal{V}$ initially has a value, $x_i(0) \in \mathbb{R}$, and the goal is for all nodes to compute the average, $\frac{1}{n} \sum_{i=1}^n x_i(0)$. In general, distributed averaging algorithms seek to achieve consensus on the average while only allowing messages to be passed between neighboring nodes, as defined by the communication graph \mathcal{G} . In broadcast gossip algorithms we make the following additional restrictions. Each node has a unique id, which corresponds to its index in the set $\{1, \dots, n\}$. When a node transmits a message, the message is received by all of its out-neighbors. The receiving nodes may know the id of the transmitter, but the transmitter will not, in general, know the ids of the receivers. Equivalently, each node knows the ids of its in-neighbors but not its out-neighbors.

2.3 Asynchronous Time Model

Following [6], we adopt the standard asynchronous time model [9]. Each node runs a clock which ticks according to an independent rate $1/n$ Poisson process. When node k 's clock ticks it initiates a broadcast gossip update, the details of which are described in the subsection that follows. Since the clocks at each node are independent, this model is equivalent to running a single, global Poisson clock which ticks at rate 1, and assigning each tick uniformly and independently to one node in \mathcal{V} . In the sequel we use the variable $t \in \{1, 2, \dots\}$ to index the ticks of this global Poisson clock. Each global clock tick corresponds to one update or iteration.

2.4 Broadcast Gossip Updates

Similar to previous broadcast gossip algorithms with companion variables [7, 13], every node i maintains two variables, $x_i(t)$ and $y_i(t)$. The first variable, $x_i(t)$, is the estimate of the average at node i after t iterations, and it is initialized to $x_i(0)$, the same initial value from Sec. 2.2. The second variable, $y_i(t)$, is the companion variable at node i after t iterations, and it is initialized to $y_i(0) = 0$. The companion variables (called ‘‘surplus’’ variables in [14]), play the role of compensating for asymmetric updates made to $x_i(t)$, and if they are updated carefully, the companion variables can be used to ensure that consensus is achieved on the average.

When a node's clock ticks, it initiates an update by broadcasting its current state and companion value. Suppose that the $t + 1$ st global clock tick occurs at node k . Then k broadcasts the values

$x_k(t)$ and $y_k(t)$, and all nodes $j \in \mathcal{N}_k^-$ which receive this information set

$$x_j(t+1) = (1 - a_{j,k})x_j(t) + a_{j,k}x_k(t) + \epsilon d_j^{(k)} y_j(t) \quad (1)$$

$$y_j(t+1) = a_{j,k}(x_j(t) - x_k(t)) + (1 - \epsilon d_j^{(k)})y_j(t) + b_{j,k}y_k(t), \quad (2)$$

where the values of the algorithm parameters $a_{j,k}$, $b_{j,k}$, $d_j^{(k)}$, and $\epsilon > 0$ will be specified below. The transmitting node k sets

$$x_k(t+1) = x_k(t) \quad (3)$$

$$y_k(t+1) = 0, \quad (4)$$

and all other nodes $i \notin k \cup \mathcal{N}_k^-$ keep

$$x_i(t+1) = x_i(t) \quad (5)$$

$$y_i(t+1) = y_i(t). \quad (6)$$

Note that the nodes need not be aware of the global clock index t to implement this protocol; they can simply update two local registers (one for x_i and one for y_i) when they either broadcast a message or receive a broadcast. Below we continue to keep track of the global clock index for the purposes of analysis.

Different choices of the parameters $a_{j,k}$, $b_{j,k}$, $d_j^{(k)}$, and ϵ lead to different broadcast gossip algorithms with different properties; we will examine two particular choices of interest in Sections 5 and 6. Note that the seminal broadcast gossip algorithm of [6] is recovered by setting $\epsilon = 0$ and $a_{j,k} = \gamma$ for all $(j, k) \in \mathcal{E}$. The broadcast gossip algorithm of [7] does not directly fit the form considered here, since in [7], the receiving nodes also use $y_k(t)$ to calculate $x_j(t+1)$.

The broadcast gossip updates (1)–(6) are linear, and below we will use tools from linear algebra, spectral graph theory, and matrix perturbation theory to analyze their convergence properties. To this end, we introduce some additional notation. Let A and B be $n \times n$ matrices with entries $[A]_{i,j} = a_{i,j}$ and $[B]_{i,j} = b_{i,j}$, respectively, satisfying

$$\begin{cases} 0 < a_{i,j} \leq 1 & \text{if } (i, j) \in \mathcal{E} \\ a_{i,j} = 0 & \text{if } (i, j) \notin \mathcal{E}, \end{cases} \quad (7)$$

and

$$\begin{cases} 0 < b_{i,j} \leq 1 & \text{if } (i, j) \in \mathcal{E} \\ b_{i,j} = 0 & \text{if } (i, j) \notin \mathcal{E}. \end{cases} \quad (8)$$

The matrices are graph-conformant in the sense that they have non-zero entries in locations corresponding to the edges of \mathcal{G} .

We write $e_k \in \mathbb{R}^n$ for the k th canonical vector—the vector with all entries equal to 0 except for the k th entry, which is equal to 1. We also write $\mathbf{1}$ (respectively $\mathbf{0}$) for a n -dimensional vector with all entries equal to 1 (respectively 0).

Define $A_k = A e_k e_k^T$ and $B_k = B e_k e_k^T$. One can verify that B_k is a $n \times n$ matrix, the k th column of B_k is identical to that of B , and all other entries of B_k are zero (and similar properties hold for A_k in relation to A). It also follows directly from the definitions of A_k and B_k that $A = \sum_{k \in \mathcal{V}} A_k$ and $B = \sum_{k \in \mathcal{V}} B_k$.

The matrices A and B can be viewed as weighted adjacency matrices of the graph \mathcal{G} (possibly assigning different weights to each edge). From this view, the matrices A_k and B_k correspond to

weighted adjacency matrices of a graph \mathcal{G}_k obtained from \mathcal{G} by eliminating all edges except those of the form (i, k) for some $i \in \mathcal{V}$, i.e., by retaining only those edges emanating from k . Thus, \mathcal{G}_k represents the graph of active edges when node k transmits.

Finally, with this view of A_k as a weighted adjacency matrix on \mathcal{G}_k , let $L_k = \text{diag}(A_k \mathbf{1}) - A_k$ denote the corresponding (directed) graph Laplacian. It follows from the definition of L_k that $L_k \mathbf{1} = \mathbf{0}$. It also follows from the definition of A_k that $\sum_{k \in \mathcal{V}} L_k = \text{diag}(A \mathbf{1}) - A \stackrel{\text{def}}{=} L$, where L is the Laplacian corresponding to the graph with weighted adjacency matrix A .

The remaining algorithm parameters to discuss are $d_j^{(k)}$ and ϵ . Let $d^{(k)} \in [0, 1]^n$ denote a vector with values satisfying

$$\begin{cases} d_j^{(k)} > 0 & \text{if } (j, k) \in \mathcal{E} \\ d_j^{(k)} = 0 & \text{if } (j, k) \notin \mathcal{E}, \end{cases} \quad (9)$$

and let $D_k = \text{diag}(d^{(k)})$ denote a diagonal matrix with $[D_k]_{i,i} = d_i^{(k)}$. The positive weights $\epsilon d_j^{(k)}$ determine the amount of j 's own surplus it injects into an update of $x_j(t+1)$ when j receives a broadcast from node k . The parameter $\epsilon > 0$ will be treated as a perturbation parameter in our analysis below, and through this analysis we will obtain: 1) an upper bound on how large ϵ can be made while still ensuring convergence, as well as 2) an indication of how ϵ affects the rate of convergence.

Define the $2n \times 2n$ matrix W_k to be

$$W_k = \begin{bmatrix} I - L_k & \epsilon D_k \\ L_k & S_k - \epsilon D_k \end{bmatrix}, \quad (10)$$

where $S_k = I - e_k e_k^T + B_k$. The general broadcast gossip updates (1)–(6) can be compactly written as

$$\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = W(t) \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, \quad (11)$$

where $W(t)$ is a random matrix with $W(t) = W_k$ when node k transmits at iteration t . In the asynchronous time model, the random matrices $W(t)$, $t = 1, \dots$, are independent and identically distributed, and $W(t) = W_k$ with probability $1/n$ for all $k \in \mathcal{V}$.

3 Convergence in Expectation

Next, we focus on identifying properties that the parameters $a_{i,j}$, $b_{i,j}$, $d_j^{(k)}$, and ϵ must satisfy in order to guarantee that the iterations (11) converge in expectation. Since L_k contains some negative entries, W_k is not nonnegative, and so standard results from nonnegative matrix analysis and the study of Markov chains are not sufficient to guarantee convergence in expectation. Our approach will make use of a combination of techniques from the theory of nonnegative matrices and perturbation theory.

Taking the conditional expectation of (11) with respect to the random node that broadcasts at each iteration, given the initial values $x(0)$ and $y(0)$, we obtain

$$\mathbb{E} \left(\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} \middle| \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \right) = \mathbb{E} \left(\prod_{t'=0}^t W(t') \right) \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \quad (12)$$

$$= \overline{W}^{t+1} \begin{bmatrix} x(0) \\ y(0) \end{bmatrix}, \quad (13)$$

where $\bar{W} = \frac{1}{n} \sum_{k \in \mathcal{V}} W_k$. One can verify that $W_k[\mathbf{1}^T \mathbf{0}^T]^T = [\mathbf{1}^T \mathbf{0}^T]^T$ for all $k \in \mathcal{V}$ since $L_k \mathbf{1} = \mathbf{0}$, and so $[\mathbf{1}^T \mathbf{0}^T]^T$ is also a right eigenvector of \bar{W} corresponding to the eigenvalue 1. The main result of this section is the following.

Theorem 1. *In addition to the constraints (7), (8), and (9) imposed on the algorithm parameters above, suppose that $\|B\|_\infty \leq 1$ or $\|B\|_1 \leq 1$. Then under the assumption that \mathcal{G} is strongly connected (Assumption 1), there exists a value $\eta > 0$ such that if $\epsilon \in (0, \eta]$, then 1 is a simple eigenvalue of \bar{W} with corresponding left eigenvector $[w_1^T \ w_2^T]^T$ normalized such that $[w_1^T \ w_2^T][\mathbf{1}^T \ \mathbf{0}^T]^T = w_1^T \mathbf{1} = 1$, and*

$$\lim_{t \rightarrow \infty} \mathbb{E} \left(\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \middle| \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \right) = \begin{bmatrix} (w_1^T x(0)) \mathbf{1} \\ \mathbf{0} \end{bmatrix}. \quad (14)$$

Remark 1. As a consequence of Theorem 1, the broadcast gossip updates (1)–(6) will converge to the average consensus if and only if $w_1 = \frac{1}{n} \mathbf{1}$. From the expression for \bar{W} derived below (see eqn. (15)), it turns out that this is only possible if $w_2 = \frac{1}{n} \mathbf{1}$ and $\mathbf{1}^T B = \mathbf{1}^T$.

The rest of this section is devoted to the proof of Theorem 1.

3.1 Preliminaries and The Plan

From (10), we find that the expected update matrix \bar{W} has the form

$$\bar{W} = \underbrace{\begin{bmatrix} I - \bar{L} & 0 \\ \bar{L} & \bar{S} \end{bmatrix}}_{\stackrel{\text{def}}{=} W_0} + \epsilon \underbrace{\begin{bmatrix} 0 & \bar{D} \\ 0 & -\bar{D} \end{bmatrix}}_{\stackrel{\text{def}}{=} E}, \quad (15)$$

where, recalling that $L = \text{diag}(A\mathbf{1}) - A$ is the Laplacian of the graph with weighted adjacency matrix A , we have

$$\bar{L} = \frac{1}{n} \sum_{k \in \mathcal{V}} L_k = \frac{1}{n} L \quad (16)$$

$$\bar{D} = \frac{1}{n} \sum_{k \in \mathcal{V}} D_k = \text{diag} \left(\sum_{k \in \mathcal{V}} d^{(k)} \right) \quad (17)$$

$$\bar{S} = \frac{1}{n} \sum_{k \in \mathcal{V}} S_k = \left(1 - \frac{1}{n} \right) I + \frac{1}{n} B. \quad (18)$$

Using the expression (15) for \bar{W} , one can verify the statement made in Remark 1 above.

From (15), it is evident that \bar{W} , can be viewed as a perturbed version of the matrix W_0 . The proof of Theorem 1 involves first characterizing the eigenvalues of W_0 using concepts from the theory of nonnegative matrices. Then results from perturbation theory can be used to determine the eigenvalues of \bar{W} as a function of ϵ and the eigenvalues of W_0 and E . Before proceeding, we briefly review background material from nonnegative matrix theory and perturbation theory.

3.2 Background

Recall that a matrix F is called *nonnegative* if all of its entries are greater than or equal to zero. A square nonnegative matrix F is *primitive* if there exists a positive integer k such that all entries of F^k are strictly positive. If F corresponds to the weighted adjacency matrix of a strongly connected graph, then it is irreducible and thus primitive [15].

Next we recall some definitions and results from perturbation theory [16].

Lemma 1 ([16] Sec. 2.4). *Suppose that a matrix $F(\epsilon)$ is continuously differentiable (entry-wise) with respect to the perturbation parameter ϵ . Then the eigenvalues of $F(\epsilon)$ are continuous functions of ϵ .*

An eigenvalue of $F(\epsilon)$ is called *stable* if it does not depend on ϵ . An eigenvalue of $F(\epsilon)$ is called *semi-simple* if its algebraic multiplicity is equal to its geometric multiplicity.

Lemma 2 ([16] Sec. 2.8). *Suppose that $F(\epsilon) = F_0 + \epsilon E$. Let λ_0 be a semi-simple double eigenvalue of W_0 with corresponding left eigenvectors u_1 and u_2 and right eigenvectors v_1 and v_2 normalized such that $u_1^T v_1 = 1$ and $u_2^T v_2 = 1$. Then, for $\epsilon > 0$, the eigenvalue λ_0 bifurcates into two distinct eigenvalues $\lambda_{0,1}(\epsilon)$ and $\lambda_{0,2}(\epsilon)$ of $F(\epsilon)$. The bifurcation is given in the form of the power series*

$$\lambda_{0,1}(\epsilon) = \lambda_0 + \epsilon \lambda' + o(\epsilon) \quad (19)$$

$$\lambda_{0,2}(\epsilon) = \lambda_0 + \epsilon \lambda'' + o(\epsilon), \quad (20)$$

where λ' and λ'' are the eigenvalues of the 2×2 matrix,

$$\begin{bmatrix} u_1^T E v_1 & u_1^T E v_2 \\ u_2^T E v_1 & u_2^T E v_2 \end{bmatrix}. \quad (21)$$

3.3 Eigenvalues of W_0

In order to apply the perturbation results mentioned above, we need to first identify the eigenvalues of W_0 . Observe, from (15), that W_0 is block diagonal, and so the eigenvalues of W_0 are the collective eigenvalues of $I - \bar{L}$ and \bar{S} .

Lemma 3. *The matrix $I - \bar{L}$ is primitive, its largest eigenvalue is 1, and all other eigenvalues of $I - \bar{L}$ have moduli strictly less than 1.*

Proof. Recall that $a_{i,j} \in [0, 1]$ according to the constraints (7), and so $I - \bar{L}$ is a square non-negative matrix. Moreover, all diagonal elements of $I - \bar{L}$ are strictly positive. Since \mathcal{G} is strongly connected, it follows that $I - \bar{L}$ is irreducible, and combining these facts gives that $I - \bar{L}$ is primitive. Since $I - \bar{L}$ is primitive and $(I - \bar{L})\mathbf{1} = \mathbf{1}$, we have $\rho(I - \bar{L}) \geq 1$. Also, its spectral radius is bounded by $\rho(I - \bar{L}) \leq \|I - \bar{L}\|_\infty = 1$. Thus its largest eigenvalue is 1 and, according to the Perron-Frobenius Theorem [17], all other eigenvalues of $I - \bar{L}$ are strictly less than 1. \square

Based on Lemma 3 we know that W_0 has at least one eigenvalue equal to 1. Next we need to determine the eigenvalues of $\bar{S} = (1 - \frac{1}{n})I + \frac{1}{n}B$. If $\lambda_i(B)$ is an eigenvalue of B , then $\lambda_i(\bar{S}) = 1 - \frac{1}{n} + \frac{1}{n}\lambda_i(B)$ is an eigenvalue of \bar{S} , and so the real task is to characterize the eigenvalues of B . If all eigenvalues of B have magnitude less than 1, then all eigenvalues of \bar{S} are also less than 1, and so 1 is a simple eigenvalue of W_0 . On the other hand, if 1 is an eigenvalue of B then it is also an eigenvalue of \bar{S} , in which case 1 is a multiple eigenvalue of W_0 .

Under the assumptions of Theorem 1 we have that $\|B\|_\infty \leq 1$ or $\|B\|_1 \leq 1$. Since $\rho(B) \leq \min\{\|B\|_\infty, \|B\|_1\}$, it follows that the largest eigenvalue of B is no larger than 1. Moreover, it follows from Assumption 1 and (8) that B corresponds to the weighted adjacency matrix of a strongly connected digraph, and hence B is primitive. Thus, $\bar{S} = (1 - \frac{1}{n})I + \frac{1}{n}B$ is also primitive and its diagonal entries are all positive. Then, from the Perron-Frobenius Theorem, the largest eigenvalue $\lambda_1(B)$ of B is simple and all other eigenvalues of B have magnitude strictly less than $\lambda_1(B)$.

It turns out that under the condition $\|B\|_\infty \leq 1$ or $\|B\|_1 \leq 1$ there are two possible cases: either $\lambda_1(\bar{S}) < 1$ or $\lambda_1(\bar{S}) = 1$. These cases are captured by the two following lemmas.

Lemma 4. *Suppose that B is either row stochastic, column stochastic, or doubly stochastic. Then 1 is a simple eigenvalue of \bar{S} and all other eigenvalues of \bar{S} have moduli strictly less than 1.*

Lemma 4 follows from standard arguments in the theory of nonnegative matrices [15, 17].

Lemma 5. *Suppose that either*

$$\max_j \sum_{i=1}^n b_{i,j} \leq 1 \quad \text{and} \quad \min_j \sum_{i=1}^n b_{i,j} < 1, \quad (22)$$

or

$$\max_i \sum_{j=1}^n b_{i,j} \leq 1 \quad \text{and} \quad \min_i \sum_{j=1}^n b_{i,j} < 1. \quad (23)$$

Then the modulus of the largest eigenvalue of \bar{S} is strictly less than 1.

Proof. We will prove the lemma for the case where (22) holds; the proof for the other case follows by a similar argument where \bar{S} is replaced with \bar{S}^T . If $\max_j \sum_{i=1}^n b_{i,j} < 1$ then the claim follows since

$$\rho(\bar{S}) \leq \|\bar{S}\|_1 = 1 - \frac{1}{n} + \frac{1}{n} \max_j \sum_{i=1}^n b_{i,j} < 1. \quad (24)$$

If

$$\min_j \sum_{i=1}^n b_{i,j} < \max_j \sum_{i=1}^n b_{i,j} = 1, \quad (25)$$

then suppose, to arrive at a contradiction, that $\rho(\bar{S}) = 1$. Let u denote the right eigenvector of \bar{S} for which $\bar{S}u = u$, and let u be normalized such that $\mathbf{1}^T u = 1$. Then $Bu = u$ also, and summing over the entries of u we find that

$$\sum_{i=1}^n u_i = \sum_{i=1}^n \left(\sum_{j=1}^n b_{i,j} u_j \right) \quad (26)$$

$$= \sum_{j=1}^n \left(\sum_{i=1}^n b_{i,j} \right) u_j \quad (27)$$

$$< \sum_{j=1}^n u_j, \quad (28)$$

where the inequality follows from the assumption (25). Since this is a contradiction, it must be true that all eigenvalues of \bar{S} are strictly less than 1 when either (22) or (23) holds. \square

To summarize, in this subsection we have shown that 1 is an eigenvalue of W_0 with multiplicity at least 1. If B is sub-stochastic (i.e., if elements of some rows and columns sum to a value less than 1), then 1 is a simple eigenvalue of W_0 . On the other hand if B is a (row-, column-, or doubly) stochastic matrix, then 1 is an eigenvalue of W_0 with multiplicity 2. In the next subsection we apply tools from perturbation theory to characterize the eigenvalues of \bar{W} .

3.4 Perturbation Analysis

The proof that the broadcast gossip iterations converge in expectation hinges on showing that 1 is a simple eigenvalue of \bar{W} with appropriate corresponding eigenvectors. When 1 is a simple eigenvalue of \bar{W} then the analysis is straightforward.

Proposition 1. *Suppose that either of the conditions (22) or (23) of Lemma 5 hold. Then there exists a number $\eta > 0$ such that if $\epsilon \in (0, \eta]$ then 1 is a simple eigenvalue of the matrix $\bar{W} = W_0 + \epsilon E$ in (15) and the moduli of all other eigenvalues of \bar{W} are strictly less than 1.*

Proof. From the discussion in Section 3.3, we know that the eigenvalues of \bar{W} are the collective eigenvalues of W_0 and \bar{S} . Under the conditions of the proposition, Lemmas 3 and 5 provide that 1 is a simple eigenvalue of W_0 and the moduli of all other eigenvalues of W_0 are strictly less than 1. Observe that $\bar{W} = W_0 + \epsilon E$ depends on ϵ in a continuous manner, and so the eigenvalues of \bar{W} are continuous functions of the perturbation parameter ϵ . Therefore, there exists $\eta > 0$ so that 1 is a simple eigenvalue of \bar{W} and the moduli of all other eigenvalues of \bar{W} are strictly less than 1 when $\epsilon \in (0, \eta]$. \square

When 1 is a double eigenvalue of \bar{W} we arrive at the same conclusion, but the proof requires a bit more effort.

Proposition 2. *Suppose that B is either row stochastic, column stochastic, or doubly stochastic. Then there exists a number $\eta > 0$ such that if $\epsilon \in (0, \eta]$ then 1 is a simple eigenvalue of the matrix \bar{W} and the moduli of all other eigenvalues of \bar{W} are strictly less than 1.*

Proof. The proof follows from a generalization of an argument in [14]. Under the conditions of the proposition, Lemmas 3 and 4 provide that 1 is an eigenvalue of W_0 with multiplicity 2. One can verify that 1 is a semi-simple eigenvalue of W_0 since there exist two linearly independent right eigenvectors u_1 and u_2 , with corresponding linearly independent left eigenvectors v_1 and v_2 . These eigenvectors are given by

$$u_1 = \begin{bmatrix} \mathbf{0} \\ \frac{1}{w^T q} q \end{bmatrix}, \quad u_2 = \begin{bmatrix} \mathbf{1} \\ \frac{-1}{w^T q} q \end{bmatrix}, \quad v_1 = \begin{bmatrix} w \\ w \end{bmatrix}, \quad v_2 = \begin{bmatrix} p \\ \mathbf{0} \end{bmatrix}, \quad (29)$$

where w is the eigenvector of \bar{S} satisfying $w^T \bar{S} = w^T$ and $w^T \mathbf{1} = 1$, and where p and q are the left and right eigenvectors of $I - \bar{L}$ corresponding to the eigenvalue 1, normalized so that $p^T \mathbf{1} = 1$ and $\mathbf{1}^T q = 1$. Note that these eigenvectors exist as a consequence of Lemmas 3 and 4. One can verify that

$$v_1^T u_1 = v_2^T u_2 = 1, \quad v_1^T u_2 = v_2^T u_1 = 0. \quad (30)$$

Also note that all entries of the vectors w , p , and q are positive since $I - \bar{L}$ and \bar{S} are primitive matrices.

Applying Lemma 2 with these values, we find that the semi-simple eigenvalue 1 bifurcates into two eigenvalues

$$\lambda_{1,1}(\epsilon) = 1 + \epsilon \lambda' + o(\epsilon) \quad (31)$$

$$\lambda_{1,2}(\epsilon) = 1 + \epsilon \lambda'' + o(\epsilon), \quad (32)$$

where λ' and λ'' are the eigenvalues of the matrix

$$\begin{bmatrix} v_1^T E u_1 & v_1^T E u_2 \\ v_2^T E u_1 & v_2^T E u_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \frac{1}{w^T q} p^T \bar{D} q & \frac{-1}{w^T q} p^T \bar{D} q \end{bmatrix}. \quad (33)$$

Thus, we have $\lambda' = 0$. Also, since p, q, w , and the diagonal entries of \overline{D} are all strictly positive, we have $\lambda'' = \frac{-1}{w^T q} p^T \overline{D} q < 0$. It follows that $\lambda_{1,1} = 1$ is the stable eigenvalue of \overline{W} corresponding to the right eigenvector $[\mathbf{1}^T \ \mathbf{0}^T]^T$. Moreover, for sufficiently small ϵ , we have $\lambda_{1,2}(\epsilon) < 1$ since $d\lambda_{1,2}(\epsilon)/d\epsilon|_{\epsilon=0} = \lambda'' < 0$. Therefore, there must exist a positive constant η_1 so that $|\lambda_{1,2}(\epsilon)| < |\lambda_{1,1}(\epsilon)|$ when $\epsilon \in (0, \eta_1]$. In addition, the eigenvalues of \overline{W} are continuous functions of ϵ , so the moduli of $\lambda_{1,1}(\epsilon) = \lambda_1(\overline{W})$ and $\lambda_{1,2}(\epsilon) = \lambda_2(\overline{W})$ will dominate the moduli of all other eigenvalues of \overline{W} provided that $\epsilon > 0$ is sufficiently small; i.e., there exists an $\eta_2 > 0$ such that $\max_{i \neq 1,2} |\lambda_i(\overline{W})| < 1$ when $\epsilon \in (0, \eta_2]$. Therefore, when $\epsilon \in (0, \min\{\eta_1, \eta_2\}]$, then 1 is a simple eigenvalue of \overline{W} and the moduli of all other eigenvalues are strictly less than 1. \square

We are now ready to complete the proof of Theorem 1.

3.5 Proof of Theorem 1

Suppose that 1 is a simple eigenvalue of a matrix \overline{W} and the moduli of all other eigenvalues are strictly less than 1. Let u and v denote the left and right eigenvectors of \overline{W} corresponding to the eigenvalue 1, normalized so that $u^T v = 1$. Then it is known [14, 18] that $\lim_{t \rightarrow \infty} \overline{W}^t = uv^T$.

From Propositions 1 and 2 we know that, under the conditions of Theorem 1, 1 is a simple eigenvalue of \overline{W} with corresponding right eigenvector $[\mathbf{1}^T \ \mathbf{0}^T]^T$, and the moduli of all other eigenvalues of \overline{W} are strictly less than 1. Let $[w_1^T \ w_2^T]^T$ be the left eigenvector of \overline{W} satisfying

$$[w_1^T \ w_2^T] \overline{W} = [w_1^T \ w_2^T], \quad (34)$$

normalized such that $[w_1^T \ w_2^T][\mathbf{1}^T \ \mathbf{0}^T]^T = 1$. Then the expected broadcast gossip updates (13) converge to a limit

$$\lim_{t \rightarrow \infty} \mathbb{E} \left(\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \middle| \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \right) = \lim_{t \rightarrow \infty} \overline{W}^t \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \quad (35)$$

$$= \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} [w_1^T \ w_2^T] \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \quad (36)$$

$$= (w_1^T x(0)) \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, \quad (37)$$

where the last line follows since $y(0) = \mathbf{0}$. This completes the proof of Theorem 1.

4 Convergence in the Second Moment

The previous section dealt with convergence in expectation. Next we present a general condition for convergence in the second moment of the broadcast gossip algorithm described in Section 2.

Theorem 2. *Suppose that B is a (row-, column-, or doubly) stochastic matrix. Let $v \in \mathbb{R}^n$ be the vector satisfying $v^T B = v^T$ normalized such that $v^T \mathbf{1} = 1$. The sequence of vectors $\{x(t), y(t)\}_{t=1}^\infty$ generated by the broadcast gossip updates (1)–(6) satisfy*

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[\left\| \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} - \begin{bmatrix} (v^T x(0)) \mathbf{1} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \middle| \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \right] = 0 \quad (38)$$

if and only if

$$\rho \left(\mathbb{E}[W(t) \otimes W(t)] - \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \right) \left(\begin{bmatrix} v \\ v \end{bmatrix} \otimes \begin{bmatrix} v \\ v \end{bmatrix} \right)^T \right) < 1. \quad (39)$$

Remark 2. Theorem 2 can be viewed as generalizing the convergence conditions for linear iterations described in [18, 19] to update matrices which have the form (10).

Proof. We first prove that (39) implies (38). Let $z(t) = [x(t)^T \ y(t)^T]^T$ and define the error vector $m(t) = z(t) - Jz(0)$, where

$$J = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} v^T & v^T \end{bmatrix}. \quad (40)$$

Observe that $W_k J = J$ for all $k \in \mathcal{V}$, since $L_k \mathbf{1} = \mathbf{0}$, and therefore $m(t+1) = W(t)m(t)$. Let $M(t) = m(t)m(t)^T$. Then $M(t+1) = W(t)M(t)W(t)^T$. Construct a vector $\tilde{m}(t) \in \mathbb{R}^{4n^2}$ by stacking the elements of $M(t)$ column-wise, and observe that

$$\mathbb{E}[\tilde{m}(t)|\tilde{m}(0)] = \prod_{s=0}^{t-1} \mathbb{E}[W(s) \otimes W(s)] \tilde{m}(0) \quad (41)$$

$$= \mathbb{E}[W(1) \otimes W(1)]^t \tilde{m}(0), \quad (42)$$

since the matrices $W(t)$ are independent and identically distributed.

Under the assumption that $v^T B = v^T$, one can verify that $[v^T \ v^T] \mathbb{E}[W(t)] = [v^T \ v^T]$. Note that such a vector exists since B is primitive* and either row or column stochastic (by assumption). We have also seen that $\mathbb{E}[W(t)][\mathbf{1}^T \ \mathbf{0}^T]^T = [\mathbf{1}^T \ \mathbf{0}^T]^T$. It follows that $([v^T \ v^T]^T \otimes [v^T \ v^T]^T)$ and $([\mathbf{1}^T \ \mathbf{0}^T]^T \otimes [\mathbf{1}^T \ \mathbf{0}^T]^T)$ are the left and right eigenvectors of $\mathbb{E}[W(t) \otimes W(t)]$ corresponding to the eigenvalue 1. If assumption (39) holds, then we have

$$\lim_{t \rightarrow \infty} \mathbb{E}[\tilde{m}(t)|\tilde{m}(0)] = \left(\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix} \right) \left(\begin{bmatrix} v \\ v \end{bmatrix} \otimes \begin{bmatrix} v \\ v \end{bmatrix} \right)^T \tilde{m}(0) \quad (43)$$

$$= 0, \quad (44)$$

where the last equality follows because v is orthogonal to $m(0)$ and hence the vector $([v^T \ v^T]^T \otimes [v^T \ v^T]^T)$ is orthogonal to $\tilde{m}(0)$.

Observe that (44) implies that $\mathbb{E}[m_i(t)^2] \rightarrow 0$ for all i , and therefore

$$\mathbb{E}[m(t)^T m(t)|m(0)] = \sum_{i=1}^{2n} \mathbb{E}[m_i(t)^2] \rightarrow 0, \quad (45)$$

which gives us (38).

Next, to see that (38) implies (39), observe that if (38) holds then it must be true that $\mathbb{E}[m_i(t)^2] \rightarrow 0$ for all $i = 1, \dots, 2n$. By the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[m_i(t)m_j(t)]^2 \leq \mathbb{E}[m_i(t)^2] \cdot \mathbb{E}[m_j(t)^2] \rightarrow 0. \quad (46)$$

Therefore, each entry in the matrix $\mathbb{E}[m(t)m(t)^T]$ tends to 0 as $t \rightarrow \infty$, independent of $m(0)$, which implies that (39) must hold. \square

5 Unbiased Broadcast Gossip

This section proposes a particular choice of values for the parameters $a_{j,k}$, $b_{j,k}$, and $d_j^{(k)}$, corresponding to a particular family of broadcast gossip algorithms. For the choice considered in this

*To see why, recall that \mathcal{G} is strongly connected (Assumption 1).

section, we guarantee that the broadcast gossip updates (1)–(6) converge to the average consensus. For this reason we refer to these as *unbiased broadcast gossip algorithms* (UBGA).

Recall that \mathcal{N}_j^- denotes the set of out-neighbors of node j and \mathcal{N}_j^+ denotes the in-neighbors of node j . Let $|\mathcal{N}|$ denote the cardinality of the set \mathcal{N} . Let $\delta_j^- = |\mathcal{N}_j^-|$ denote the out-degree of node j , and let $\delta_j^+ = |\mathcal{N}_j^+|$ denote its in-degree.

Unbiased broadcast gossip algorithms are obtained by setting

$$d_j^{(k)} = \begin{cases} 1/\delta_j^+ & \text{if } j \in \mathcal{N}_k^- \\ 0 & \text{otherwise} \end{cases} \quad (47)$$

$$B_{j,k} = \begin{cases} 1/\delta_k^- & \text{if } j \in \mathcal{N}_k^- \\ 0 & \text{otherwise,} \end{cases} \quad (48)$$

and taking $A_{j,k}$ to be any values which satisfy the constraints (7). In order to implement such a protocol, each node $j \in \mathcal{N}_k^-$ that receives messages from k needs to know δ_k^- , the out-degree of the broadcasting node k . If k knows its out-degree (i.e., the number of neighbors that receive its broadcasts) then this can be accomplished by having k broadcast the value of δ_k^- to all nodes in \mathcal{N}_k^- . If \mathcal{G} is undirected[†], as is assumed in [6], then $\delta_k^- = \delta_k^+$, and so it is reasonable for k to know its out-degree. On the other hand, in a general directed graph it may be difficult or impractical for k to know its out degree since k may not receive messages directly from all nodes $j \in \mathcal{N}_k^-$. In Section 6 below we describe and analyze an alternative algorithm which does not require knowledge of δ_k^- , but for which we are not guaranteed to achieve consensus on the average. First we discuss theoretical convergence guarantees for UBGA.

First, note that for ϵ sufficiently small, UBGA asymptotically converges in expectation to the average consensus in the sense that $\mathbb{E}[x(t)] \rightarrow \frac{1}{n}\mathbf{1}^T x(0)$ and $\mathbb{E}[y(t)] \rightarrow \mathbf{0}$. To see why, observe that for the choice of parameters given in (48) we have $\mathbf{1}^T B = \mathbf{1}^T$. Therefore, by Theorem 1 (see also Remark 1), there exists an $\eta > 0$ such that UBGA converges in expectation to the average consensus for $\epsilon \in (0, \eta]$. It turns out that UBGA also converges to the average consensus solution in the mean-squared sense.

Proposition 3. *Let the parameters $d_j^{(k)}$ and $B_{j,k}$ be chosen as in (47) and (48), and take $v = \frac{1}{n}\mathbf{1}$. Then there exists a constant $\eta > 0$ such that if $\epsilon \in (0, \eta]$ then (39) holds, and so*

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[\left\| \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} - \begin{bmatrix} \frac{1}{n} (\mathbf{1}^T x(0)) \mathbf{1} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 \mid \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} \right] = 0. \quad (49)$$

Proof. To prove the theorem, we need to show that there exists an $\eta > 0$ such that (39) holds. Convergence in the second moment then follows from Theorem 2.

To show that (39) holds for sufficiently small ϵ , we use a perturbation argument similar to one in [14]. Since $(I - \mathbb{E}[L_k])\mathbf{1} = \mathbf{1}$, the matrix $I - \mathbb{E}[L_k]$ has 1 as an eigenvalue with right eigenvector $\mathbf{1}$. It follows that there exists a corresponding left eigenvector w satisfying $w^T(I - \mathbb{E}[L_k]) = w^T$ and $w^T \mathbf{1} = 1$, and all entries of w are positive. Consequently, the following four equalities hold:

$$\mathbb{E}[(I - L_k) \otimes (I - L_k)] (\mathbf{1} \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1}) \quad (50)$$

$$(w \otimes \mathbf{1})^T \mathbb{E}[(I - L_k) \otimes S_k] = (w \otimes \mathbf{1})^T \quad (51)$$

$$(\mathbf{1} \otimes w)^T \mathbb{E}[S_k \otimes (I - L_k)] = (\mathbf{1} \otimes w)^T \quad (52)$$

$$(\mathbf{1} \otimes \mathbf{1})^T \mathbb{E}[S_k \otimes S_k] = (\mathbf{1} \otimes \mathbf{1})^T. \quad (53)$$

[†]I.e., $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$

Thus, the four matrices $\mathbb{E}[(I - L_k) \otimes (I - L_k)]$, $\mathbb{E}[(I - L_k) \otimes S_k]$, $\mathbb{E}[S_k \otimes (I - L_k)]$, and $\mathbb{E}[S_k \otimes S_k]$ have eigenvalue 1. These matrices are all non-negative and irreducible under the constraints (7) and since \mathcal{G} is strongly connected. Moreover, since the corresponding eigenvectors $(\mathbf{1} \otimes \mathbf{1})$, $(w \otimes \mathbf{1})^T$, $(\mathbf{1} \otimes w)^T$, and $(\mathbf{1} \otimes \mathbf{1})^T$ are all positive, it follows from the Perron-Frobenius Theorem that 1 is the largest eigenvalue of the four matrices above, and all other eigenvalues have moduli strictly less than 1.

In order to apply a perturbation argument to analyze the eigenvalues of $\mathbb{E}[W(t) \otimes W(t)]$, let us write $W_k = M_k + \epsilon F_k$ with

$$M_k = \begin{bmatrix} I - L_k & 0 \\ L_k & S_k \end{bmatrix} \quad (54)$$

$$F_k = \begin{bmatrix} 0 & D_k \\ 0 & -D_k \end{bmatrix}. \quad (55)$$

Similarly, we have $W(t) = M(t) + \epsilon F(t)$, with $M(t)$ and $F(t)$ being random matrices drawn from the collection $\{(M_k, F_k)\}$ depending on which node k broadcasts at iteration t . In the following we omit the dependence on t to simplify notation. With the above definitions, we have $\mathbb{E}[W \otimes W] = \mathbb{E}[M \otimes M] + \epsilon \mathbb{E}[M \otimes F + F \otimes M + F \otimes F]$. There is a $2n \times 2n$ permutation matrix P such that $P^T \mathbb{E}[W \otimes W] P = \widehat{M} + \epsilon \widehat{F}$ where

$$\widehat{M} = \mathbb{E} \begin{bmatrix} (I - L) \otimes (I - L) & 0 & 0 & 0 \\ (I - L) \otimes L & (I - L) \otimes S & 0 & 0 \\ L \otimes (I - L) & 0 & S \otimes (I - L) & 0 \\ L \otimes L & L \otimes S & S \otimes L & S \otimes S \end{bmatrix} \quad (56)$$

$$\widehat{F} = \mathbb{E} \begin{bmatrix} 0 & (I - L) \otimes D & D \otimes (I - L) & D \otimes \epsilon D \\ 0 & -(I - L) \otimes D & D \otimes L & D \otimes (S - \epsilon D) \\ 0 & L \otimes D & -D \otimes (I - L) & (S - \epsilon D) \otimes D \\ 0 & -L \otimes D & -D \otimes L & -S \otimes D - D \otimes S + D \otimes \epsilon D \end{bmatrix}. \quad (57)$$

Since 1 is a simple eigenvalue of each of the matrices $\mathbb{E}[(I - L) \otimes (I - L)]$, $\mathbb{E}[(I - L) \otimes S]$, $\mathbb{E}[S \otimes (I - L)]$, and $\mathbb{E}[S \otimes S]$, and all other eigenvalues of these matrices have moduli less than 1, we find that 1 is a semi-simple eigenvalue of \widehat{M} with multiplicity 4, and all other eigenvalues of \widehat{M} are strictly less than 1.

We will use a perturbation argument to show that for sufficiently small $\epsilon > 0$, the largest eigenvalue of $\widehat{M} + \epsilon \widehat{F}$ is 1 and 1 is a simple eigenvalue. Our argument is based on a generalization of Lemma 2 that addresses bifurcation of a quadruple semi-simple eigenvalues rather than double semi-simple eigenvalues [16].

Let $\widehat{\lambda}_i(\epsilon)$, $i = 1, \dots, 4$ denote the four bifurcating eigenvalues of $\widehat{M} + \epsilon \widehat{F}$. Similar to Lemma 2, we have $\widehat{\lambda}_i(\epsilon) = 1 + \epsilon \xi_i + o(\epsilon)$, where ξ_i , $i = 1, \dots, 4$ are four eigenvalues of the matrix of similar structure to (21). Solving for ξ_1, \dots, ξ_4 , we find that the derivatives of the eigenvalues $\widehat{\lambda}_i(\epsilon)$ with respect to ϵ are given by

$$\frac{d\widehat{\lambda}_1(\epsilon)}{d\epsilon} = 0 \quad (58)$$

$$\frac{d\widehat{\lambda}_2(\epsilon)}{d\epsilon} = \frac{d\widehat{\lambda}_3(\epsilon)}{d\epsilon} = -N v_1^T \mathbb{E}[D] v_2 < 0 \quad (59)$$

$$\frac{d\widehat{\lambda}_4(\epsilon)}{d\epsilon} = -2N v_1^T \mathbb{E}[D] v_2 < 0, \quad (60)$$

where v_1 is the positive left eigenvector of $I - L$ normalized such that $v^T \mathbf{1} = 1$, and v_2 is the positive right eigenvector of S normalized so that $\mathbf{1}^T v_2 = 1$. Similar as in Lemma 2 (see [16]), it follows that there exists a real number $\eta > 0$ such that the matrix $\mathbb{E}[W \otimes W]$ has only one simple eigenvalue 1, and the moduli of all other eigenvalues are smaller than 1 when $\epsilon \in (0, \eta]$. Thus, (39) holds, and convergence in the second moment follows from Theorem 2. \square

6 Biased Broadcast Gossip

The previous section proposed UBGA, a broadcast gossip algorithm which provably converges to the average consensus in both expectation and in the mean-squared sense. UBGA is practical in situations when the network can be guaranteed to be undirected, or when nodes otherwise know their out-degree. For instance, one could enforce that only symmetric links are used by having each node broadcast its set of in-neighbors and then only updating using messages from neighbors for which the neighborhood relationship is symmetric. However, this may be undesirable in some applications, and so in this section we consider an alternative family of broadcast gossip algorithms. These algorithms are no longer guaranteed to converge to an average consensus, and so we refer to them as *biased broadcast gossip algorithms* (BBGAs). However, we still guarantee convergence in expectation and in the mean-squared sense to a characterizable value which depends on the initial state at each node and the structure of the network.

Biased broadcast gossip algorithms are obtained by setting

$$d_j^{(k)} = \begin{cases} 1/\delta_j^+ & \text{if } j \in \mathcal{N}_k^- \\ 0 & \text{otherwise} \end{cases} \quad (61)$$

$$B_{j,k} = \begin{cases} 1/\delta_j^+ & \text{if } j \in \mathcal{N}_k^- \\ 0 & \text{otherwise,} \end{cases} \quad (62)$$

and taking $A_{j,k}$ to be any values which satisfy the constraints (7). To implement such a scheme we only require that each node has knowledge of its in-neighbors, which is reasonable in the broadcast setting.

Observe that, for the choice of parameters just specified, both $I - L_k$ and \bar{S} are row-stochastic matrices. Let v be such that $v^T B = v^T$ and $v^T \mathbf{1} = 1$. Thus, the entries of v satisfy $v_k = \sum_{j \in \mathcal{N}_k^-} v_j / \delta_j^+$, and all entries of v are positive. Such an eigenvector exists since B is also row-stochastic. One can verify that $v^T S_k = v^T$ also holds, and so $v^T \bar{S} = v^T$. In general, we do not have $v = \frac{1}{n} \mathbf{1}$ unless $\delta_j^+ = \delta_j^-$ for all j and $\delta_j^+ = \delta_i^+$ for all $i \neq j$. Therefore convergence to the average consensus can no longer be guaranteed in general. However, we still obtain convergence in expectation to a (non-average) consensus, via Theorem 1, and we can also show that BBGA converges in the second moment.

Proposition 4. *Let the parameters $d_j^{(k)}$ and $B_{j,k}$ be chosen as in (61) and (62). There exists $\eta > 0$ such that if $\epsilon \in (0, \eta]$ then (39) holds and so (38) also holds with v being the vector such that $v^T B = v^T$ and $v^T \mathbf{1} = 1$.*

Proof. To prove the claim we show that (39) holds and then invoke Theorem 2. We use an argument similar to that used in the proof of Proposition 3. Since $(I - L_k)\mathbf{1}$, $S_k\mathbf{1} = \mathbf{1}$, and $v^T S_k = v^T$, the

following four equalities hold:

$$\mathbb{E}[(I - L_k) \otimes (I - L_k)](\mathbf{1} \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1}) \quad (63)$$

$$\mathbb{E}[(I - L_k) \otimes S_k](\mathbf{1} \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1}) \quad (64)$$

$$\mathbb{E}[S_k \otimes (I - L_k)](\mathbf{1} \otimes \mathbf{1}) = (\mathbf{1} \otimes \mathbf{1}) \quad (65)$$

$$(v^T \otimes v^T)\mathbb{E}[S_k \otimes S_k] = (v^T \otimes v^T). \quad (66)$$

Since the eigenvectors above are all positive, it follows that 1 is the largest eigenvalue of each of the four matrices $\mathbb{E}[(I - L_k) \otimes (I - L_k)]$, $\mathbb{E}[(I - L_k) \otimes S_k]$, $\mathbb{E}[S_k \otimes (I - L_k)]$, and $\mathbb{E}[S_k \otimes S_k]$.

Similar to the proof of Proposition 3, we find that \widehat{M} has largest eigenvalue 1 with multiplicity 4, and all the moduli of all other eigenvalues are strictly smaller than 1. Let $\lambda_1(\epsilon), \dots, \lambda_4(\epsilon)$ denote the four corresponding eigenvalues of $\mathbb{E}[W \otimes W]$. In this case, we can solve for the eigenvalues and again take their derivatives to find that

$$\frac{d\widehat{\lambda}_1(\epsilon)}{d\epsilon} = 0 \quad (67)$$

$$\frac{d\widehat{\lambda}_2(\epsilon)}{d\epsilon} = \frac{d\widehat{\lambda}_3(\epsilon)}{d\epsilon} = -v_1^T \mathbb{E}[D] \mathbf{1} < 0 \quad (68)$$

$$\frac{d\widehat{\lambda}_4(\epsilon)}{d\epsilon} = -2v_1^T \mathbb{E}[D] \mathbf{1} < 0, \quad (69)$$

where v_1 satisfies $v_1^T \mathbb{E}[I - L_k] = v_1^T$ and $v_1^T \mathbf{1} = 1$. Thus, there exists a positive scalar $\eta > 0$ such that 1 is a simple eigenvalue of $\mathbb{E}[W \otimes W]$ and all other eigenvalues are strictly less than 1 when $\epsilon \in (0, \eta]$. Subsequently, (39) holds, and convergence in the second moment follows from Theorem 2. \square

7 Upper Bound on η

So far we have demonstrated that there exist broadcast gossip algorithms of the form described in Section 2 which are guaranteed to converge when the parameter ϵ is chosen to be sufficiently small. In this section we derive bounds on η which can be used as practical guidelines for setting this parameter. Previous results suggest that, in general, one must take $\eta = \Theta(n^{-n})$, which is extremely conservative [14, 20, 21]. The bounds in this section make use of the specific structure of $W(t)$ to obtain tighter, more useful bounds.

We begin with a simple observation related to the expected BBGA update matrix.

Lemma 6. *For updates using the BBGA parameters and for sufficiently small $\epsilon > 0$, the second largest eigenvalue of \overline{W} is $1 - \epsilon/n$.*

Proof. One can verify that, for BBGA, $\overline{W}[\mathbf{1}^T \mathbf{0}^T]^T = [\mathbf{1}^T \mathbf{0}^T]^T$ and $\overline{W}[\mathbf{1}^T - \mathbf{1}^T]^T = (1 - \epsilon/n)[\mathbf{1}^T - \mathbf{1}^T]^T$. Thus, 1 and $1 - \epsilon/n$ are eigenvalues of \overline{W} . According to Lemma 1, the eigenvalues of \overline{W} are continuous functions of ϵ . If 1 or $1 - \epsilon/n$ are not eigenvalues stemming from the semi-simple double eigenvalue 1 of W_0 , then we obtain a contradiction as $\epsilon \rightarrow 0$. Therefore, $1 - \epsilon/n$ must be the second largest eigenvalue of \overline{W} for sufficiently small ϵ . \square

In order to provide a tight characterization of the upper bound η on the perturbation parameter, we need to make more specific assumptions about the values of the weights $a_{j,k}$. Previously, we only

assumed that they satisfy the constraints (7). For the remainder of the article, unless otherwise stated, we assume that

$$a_{j,k} = \begin{cases} 1/\delta_j^+ & \text{if } j \in \mathcal{N}_k^- \\ 0 & \text{otherwise.} \end{cases} \quad (70)$$

In this case, for BBGA, we have $\bar{S} = I - \bar{L}$.

Let ξ_i denote the eigenvalues of the graph Laplacian $L = \text{diag}(A\mathbf{1}) - A$ sorted by increasing real part [22], where n is counted with multiplicity; i.e., $0 = \text{Re}(\xi_1) \leq \text{Re}(\xi_2) \leq \dots \leq \text{Re}(\xi_n) \leq 2$. Note that this last inequality holds because $\rho(L) \leq \|L\|_\infty = 2$, and the real part of ξ_k for any k is nonnegative because L is diagonally dominant. We have the following lemma characterizing the eigenvalues of \bar{W} for BBGA.

Lemma 7. *For BBGA, the $2n$ eigenvalues of \bar{W} are*

$$\lambda_{k,1} = 1 - \frac{1}{n}\xi_k - \frac{\epsilon}{2n} - \frac{1}{n}\sqrt{\epsilon\xi_k + \frac{\epsilon^2}{4}} \quad \text{for } k = 1, \dots, n, \quad (71)$$

$$\lambda_{k,2} = 1 - \frac{1}{n}\xi_k - \frac{\epsilon}{2n} + \frac{1}{n}\sqrt{\epsilon\xi_k + \frac{\epsilon^2}{4}} \quad \text{for } k = 1, \dots, n. \quad (72)$$

Proof. Observe that (61) implies $\bar{D} = \frac{1}{n}I$ for BBGA. The characteristic polynomial of \bar{W} is

$$\det(\lambda I - \bar{W}) = \det \left(\begin{bmatrix} (\lambda - 1)I + \frac{1}{n}L & -\frac{\epsilon}{n}I \\ -\frac{1}{n}L & (\lambda - 1 + \frac{\epsilon}{n})I + \frac{1}{n}L \end{bmatrix} \right) \quad (73)$$

$$= \det \left(\left[(\lambda - 1) \left(\lambda - 1 + \frac{\epsilon}{n} \right) I + \frac{2(\lambda - 1)}{n}L + \frac{1}{n^2}L^2 \right] \right), \quad (74)$$

where the last inequality follows since the matrices I and L , as well as any linear combination of them, are commutative. The zeros of the characteristic polynomial of \bar{W} thus correspond to zero eigenvalues of the matrix

$$U = (\lambda - 1) \left(\lambda - 1 + \frac{\epsilon}{n} \right) I + \frac{2(\lambda - 1)}{n}L + \frac{1}{n^2}L^2. \quad (75)$$

According to the spectral mapping theorem [15], for any eigenvalue ξ_i of the matrix L , the matrix U has a corresponding eigenvalue $(\lambda - 1)(\lambda - 1 + \frac{\epsilon}{n}) + \frac{2(\lambda - 1)}{n}\xi_i + \frac{1}{n^2}\xi_i^2$. Therefore, given an eigenvalue λ_k of \bar{W} , there must exist an eigenvalue ξ_j of L such that

$$(\lambda_k - 1) \left(\lambda_k - 1 + \frac{\epsilon}{n} \right) + \frac{2(\lambda_k - 1)}{n}\xi_j + \frac{1}{n^2}\xi_j^2 = 0. \quad (76)$$

According to Lemma 1, λ_k is a continuous function of ϵ , and so as $\epsilon \rightarrow 0$, λ_k is an eigenvalue of W_0 . For the particular choice of parameters made above, we have that $I - \bar{L} = \bar{S}$. Since ξ_k is an eigenvalue of L , we have that $1 - \frac{1}{n}\xi_k$ is an eigenvalue of $I - \bar{L} = \bar{S} = I - \frac{1}{n}L$. Thus, in the continuous limit as $\epsilon \rightarrow 0$, we have that $\lambda_k = 1 - \frac{1}{n}\xi_k$ is an eigenvalue of \bar{W} . Taking $\epsilon = 0$ and substituting for λ_k in (76), we find that $(\xi_j - \xi_k)^2 = 0$, which implies that $j = k$. In this case, (76) can be simplified to

$$(\lambda_k - 1) \left(\lambda_k - 1 + \frac{\epsilon}{n} \right) + \frac{2(\lambda_k - 1)}{n}\xi_k + \frac{1}{n^2}\xi_k^2 = 0. \quad (77)$$

Solving for the roots of these quadratic polynomials in λ_k proves the claim. \square

We have already seen that L has an eigenvalue $\lambda_1 = 0$. From Lemma 7, we again find that $\lambda_{1,1} = 1 - \epsilon/n$ and $\lambda_{1,2} = 1$ are eigenvalues of \overline{W} . If all eigenvalues of L are real, then all eigenvalues of \overline{W} are also real. In this case, we can use the monotonic ordering of eigenvalues to determine an upper bound η on ϵ to ensure that BBGA converges in expectation. We restrict to the case where all eigenvalues of L are real. When L has some complex eigenvalues monotonicity is no longer preserved making it difficult to determine a reasonable bound. The main result of this section is as follows.

Proposition 5. *Consider BBGA updates and suppose that all eigenvalues of L are real. Then BBGA converges in expectation when $0 \leq \epsilon \leq 2n + \xi_n^2/(2n) - 2\xi_n$, where ξ_n is the largest eigenvalue of L .*

The proof of Proposition 5 relies on two intermediate lemmas.

Lemma 8. *For BBGA updates, if all eigenvalues of L are real, then there is one stable eigenvalue $\lambda_{1,2} = 1$. All other eigenvalues $\lambda_{k,1}$ are strictly decreasing functions of ϵ , and all eigenvalues $\lambda_{k,2}$ are strictly increasing functions of ϵ .*

Proof. To prove the claim, differentiate (71) and (72) with respect to ϵ :

$$\frac{d\lambda_{k,1}}{d\epsilon} = -\frac{1}{2n} \left(1 + \frac{2\xi_k + \epsilon}{\sqrt{4\epsilon\xi_k + \epsilon^2}} \right) \quad (78)$$

$$\leq -\frac{1}{2n} \left(1 + \frac{2\xi_k + \epsilon}{\sqrt{(2\xi_k + \epsilon)^2}} \right) \quad (79)$$

$$= -\frac{1}{n} \quad (80)$$

$$< 0, \quad (81)$$

and

$$\frac{d\lambda_{k,2}}{d\epsilon} = -\frac{1}{2n} \left(1 - \frac{2\xi_k + \epsilon}{\sqrt{4\epsilon\xi_k + \epsilon^2}} \right) \quad (82)$$

$$\geq -\frac{1}{2n} \left(1 - \frac{2\xi_k - \epsilon}{\sqrt{(2\xi_k + \epsilon)^2}} \right) \quad (83)$$

$$= 0. \quad (84)$$

Note that equality holds in (83) if and only if $\xi_k = 0$. \square

Lemma 9. *For BBGA updates, If all eigenvalues of L are real, then the eigenvalues $\lambda_{k,1}$ and $\lambda_{k,2}$ are monotonic decreasing functions of ξ_k .*

Proof. It is clear from (71) that $\lambda_{k,1}$ is a monotonic decreasing function of ξ_k . For $\lambda_{k,2}$ and for fixed $\epsilon > 0$, observe that

$$\frac{d\lambda_{k,2}}{d\xi_k} = -\frac{1}{n} + \frac{1}{n\sqrt{4\xi_k/\epsilon + 1}} \leq 0, \quad (85)$$

where the inequality follows since $\xi_k \geq 0$. Therefore, $\lambda_{k,2}$ is also a monotonic decreasing function of ξ_k . \square

Proof of Proposition 5. Under the assumption that the eigenvalues of L are nonnegative real numbers, we have

$$\frac{\epsilon}{2} \leq \sqrt{\epsilon \xi_k + \frac{\epsilon^2}{4}} \leq \sqrt{\left(\xi_k + \frac{\epsilon}{2}\right)^2} = \frac{\epsilon}{2} + \xi_k. \quad (86)$$

Substituting (86) into (71) and (72), we get

$$1 - \frac{\epsilon}{n} - \frac{2\xi_k}{n} \leq \lambda_{k,1} \leq 1 - \frac{\epsilon}{n} - \frac{\xi_k}{n}, \quad (87)$$

$$1 - \frac{\xi_k}{n} \leq \lambda_{k,2} \leq 1, \quad (88)$$

where the equalities hold if and only if $k = 1$ with corresponding $\xi_k = 0$. From these expressions, it is clear that $\lambda_{1,2} = 1$ is a simple eigenvalue of \overline{W} and all other eigenvalues are strictly smaller than 1 since $0 \leq \xi_k \leq 2$ and $\epsilon > 0$. Furthermore, convergence is guaranteed when all eigenvalues are strictly larger than -1 . For a given ξ_k , observe that $\lambda_{k,1} \leq \lambda_{k,2}$. In addition, from Lemma 9, we have that $\lambda_{k,1}$ is a monotonic decreasing function of ξ_k . Therefore, $\lambda_{k,2} \geq \lambda_{k,1} \geq \lambda_{n,1}$ for all $k = 1, \dots, n$. Thus, we focus on determining conditions under which $\lambda_{n,1} > -1$. According to Lemma 8, $\lambda_{n,1}$ is a strictly decreasing function of ϵ . If there exists η such that $\lambda_{n,1} = -1$ when $\epsilon = \eta$, then $\lambda_{n,1} > -1$ when $\epsilon < \eta$. Solving (71) for $\lambda_{n,1} = -1$ we obtain

$$\eta = 2n + \frac{\xi_n^2}{2n} - 2\xi_n, \quad (89)$$

which completes the proof. \square

Remark 3. In general, the value of ξ_n depends on the network topology, and it may not be easy to determine a precise value of ξ_n . A more practical guideline is to take $\epsilon \in (0, \frac{2}{n}(n-1)^2)$. To see why this is reasonable, differentiate (89) with respect to ξ_n :

$$\frac{d\eta}{d\xi_n} = \frac{\xi_n}{n} - 2 \leq \frac{2}{n} - 2 \leq 0. \quad (90)$$

Therefore, η is a monotonic decreasing function of ξ_n , and η thus satisfies $\frac{2}{n}(n-1)^2 \leq \eta \leq 2n$ since $0 \leq \xi_n \leq 2$. If the perturbation parameter ϵ is not larger than $\frac{2}{n}(n-1)^2$ then BBGA is guaranteed to converge in expectation.

Note that, from Remark 3, the upper bound η is not smaller than 1. In the following section we investigate what value of ϵ leads to the fastest convergence. We find that we typically seek values of ϵ less than 1, and so this upper bound will suffice.

Although the guidelines derived above are for BBGA, in extensive simulations we have observed that the maximal value of ϵ under which UBGA still converges is typically no different than that for BBGA for a given graph. Therefore, the guidelines derived above can be also used as approximate guidelines for setting the parameters of UBGA.

8 Optimal Perturbation Parameter

In the previous section we determined an upper bound on the perturbation parameter ϵ under which convergence in expectation is guaranteed. This can be viewed as a sort of stability result. In this section we investigate what value of the perturbation parameter leads to the fastest rate of convergence. It is well known that the worst-case rate of convergence of systems of the form (13)

is governed by the second largest eigenvalue of \bar{W} . In the previous section we saw that, for BBGA, this second largest eigenvalue is $1 - \epsilon/n$ if the perturbation parameter $\epsilon > 0$ is sufficiently small, and this eigenvalue is a monotonic decreasing function of ϵ . At the same time, other eigenvalues of \bar{W} are monotonic increasing, and so it follows that the optimal value of ϵ is the one where the modulus of $1 - \epsilon/n$ first coincides with the modulus of another eigenvalue of \bar{W} .

Theorem 3. *Consider the expected update matrix \bar{W} corresponding to BBGA, and suppose that all eigenvalues of the Laplacian $L = \text{diag}(A\mathbf{1}) - A$ are real. For networks with at least three nodes, the modulus of the second largest eigenvalue of \bar{W} is minimized when the perturbation parameter is equal to $\epsilon^* = \xi_2/2$, where ξ_2 is the second largest eigenvalue of the graph Laplacian L . In this case, the second largest eigenvalue of \bar{W} is $1 - \xi_2/(2n)$. When $n = 2$, the second largest eigenvalue of \bar{W} is minimized by $\epsilon^* = 2 - \sqrt{2}$.*

Proof. According to Lemma 8, for a fixed $\epsilon > 0$, the eigenvalues of \bar{W} satisfy

$$1 - \epsilon/n = \lambda_{1,1} \geq \lambda_{2,1} \geq \dots \geq \lambda_{n,1}, \quad (91)$$

and

$$1 = \lambda_{1,2} \geq \lambda_{2,2} \geq \dots \geq \lambda_{n,2}. \quad (92)$$

We also have monotonicity of the respective eigenvalues as a function of ϵ from Lemma 9. Because the eigenvalues are continuous functions of ϵ , it follows that there are two points of interest where the second largest eigenvalue (in modulus) may switch from being $\lambda_{1,1} = 1 - \epsilon/n$. These are the points ϵ_1 where $\lambda_{1,1}(\epsilon_1) = \lambda_{2,2}(\epsilon_1)$ and ϵ_2 where $\lambda_{1,1}(\epsilon_2) = -\lambda_{n,1}(\epsilon_2)$. To complete the proof we can solve for these two values of ϵ and then determine that $\epsilon^* = \min\{\epsilon_1, \epsilon_2\}$.

Solving $\lambda_{1,1}(\epsilon_1) = \lambda_{2,2}(\epsilon_1)$, we have $\epsilon_1 = \xi_2/2$ and the corresponding eigenvalue of \bar{W} is $1 - \xi_2/(2n)$.

To solve $\lambda_{1,1}(\epsilon_2) = -\lambda_{n,1}(\epsilon_2)$, observe that

$$\epsilon_2 = \frac{3n - \xi_n - \sqrt{n^2 + 2n\xi_n - \xi_n^2}}{2} \quad (93)$$

$$\geq \frac{3n - \xi_n - \sqrt{(n + \xi_n)^2}}{2} \quad (94)$$

$$= n - \xi_n. \quad (95)$$

Since $0 \leq \xi_n \leq 2$, it must be that $\epsilon_2 > 1$ if $n \geq 3$, from which we find that $\epsilon_2 > \epsilon_1$ since $\xi_2/2 < 1$. Therefore, the optimal perturbation parameter is $\epsilon^* = \epsilon_1 = \xi_2/2$ when $n \geq 3$. If $n = 2$, then there is only one non-zero eigenvalue of the weighted Laplacian matrix L and it is equal to 2. In this case, $\epsilon_2 = 2 - \sqrt{2} < 1 = \epsilon_1$, so the optimal perturbation parameter is $\epsilon^* = 2 - \sqrt{2}$. \square

Remark 4. Note that since the modulus of the second largest eigenvalue of \bar{W} satisfies $|\lambda_{1,1}(\epsilon^*)| = \lambda_{1,1}(\epsilon^*) < \lambda_{1,1}(0) < 1$, we see that BBGA is guaranteed to converge in expectation for this setting.

The above analysis focused on the case where the eigenvalues of L are assumed to be real. In extensive simulations, we have observed that this is the case whenever \mathcal{G} is undirected, regardless of whether the edge weights are symmetric. For digraphs, the eigenvalues ξ_i of L are generally complex numbers, so a monotonicity property such as that obtained in Lemma 9 is no longer readily available. Below we analyze the optimal value of the perturbation parameter on random digraphs via simulation. We find that $\tilde{\epsilon} = \text{Re}(\xi_2)/2$ is a good guideline for directed graphs.

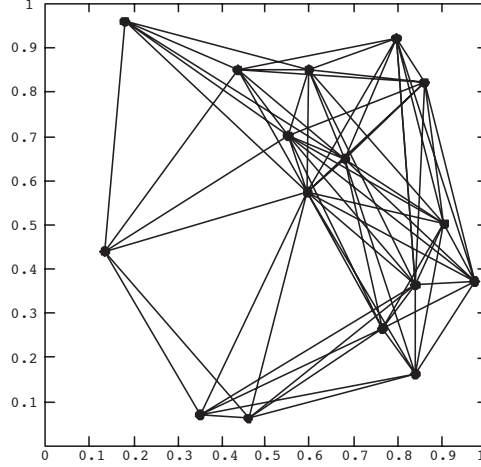


Figure 1: An example for undirected graphs with 16 nodes

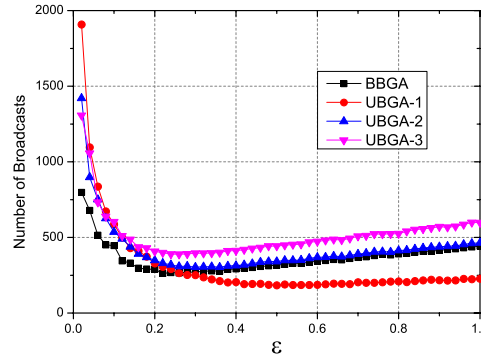


Figure 2: Number of broadcasts to converge with respect to ϵ for simulations on the graph shown in Fig. 1.

8.1 Undirected Graphs

Consider an undirected graph as illustrated in Figure 1 with 16 nodes distributed uniformly in the unit square. Nodes are connected if the Euclidean distance between them is no more than $\sqrt{2 \log n / n}$ so that the graph \mathcal{G} is connected with probability at least $1 - 1/n^2$ [23, 24]; this is the standard random geometric graph model. For the graph shown in Fig. 1, the second smallest eigenvalue of weighted Laplacian matrix L for BBGA is $\xi_2 = 0.5335$ so the optimal perturbation parameter for BBGA is $\epsilon^* = \xi_2/2 = 0.2668$.

Figure 2 shows the number of broadcasts to achieve consensus as a function of ϵ . Each point is an average over 100 trials, and we sweep over values of ϵ from 0.02 to 1 in increments of 0.02. The initial values of all nodes are independent and uniformly distributed between 0 and 1. Recall the error vector $m(t)$ defined in the proof of Theorem 2. We declare that consensus is achieved at the first iteration t where $\|m(t) - m(t-1)\|_2 \leq 10^{-5}$. Here we compare four broadcasts gossip algorithms. For BBGA we use the weights $a_{j,k}$ as defined in (70). The three versions of UBGA

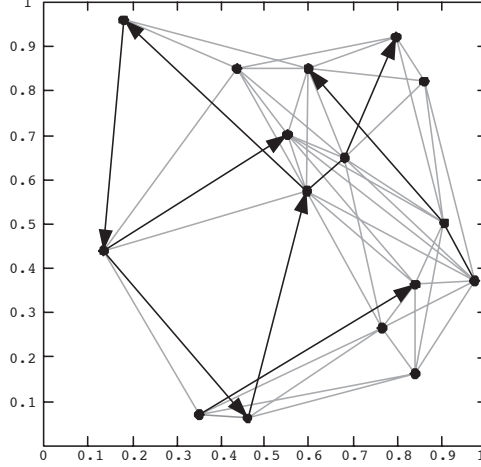


Figure 3: An example for digraphs with 16 nodes. The gray lines denote undirected edges and lines with one arrow denote directed edges.

have different choices of weights $a_{j,k}$; they are

$$a_{j,k} = \begin{cases} 0.5 & \text{if } j \in \mathcal{N}_k^- \text{ for UBGA-1,} \\ 1/\delta_j^+ & \text{if } j \in \mathcal{N}_k^- \text{ for UBGA-2,} \\ 1/\delta_j^- & \text{if } j \in \mathcal{N}_k^- \text{ for UBGA-3,} \\ 0 & \text{if } j \notin \mathcal{N}_k^-. \end{cases} \quad (96)$$

Observe that the fastest convergence for BBGA occurs near $\epsilon = 0.26$, which matches the value predicted by Proposition 3. Also observe that all three versions of UBGA have larger optimal perturbation parameter than BBGA. Using any version of UBGA at the optimal value ϵ^* for BBGA results in suboptimal performance. UBGA-1 exhibits a number advantages over the other algorithms: it converges in fewer broadcasts than the other algorithms for suitably chosen ϵ , and the curve for UBGA-1 in Fig. 2 is extremely flat near the optimal value, so its performance is very robust to the choice of ϵ in this region. From a practical perspective, UBGA-1 is also easy to implement in undirected networks since all weights $a_{j,k}$ are constants only depending on the network connectivity. For the graph shown in Fig. 1, the largest eigenvalue of L for BBGA is 1.3796, and the corresponding upper bound for ϵ is 29.30, which can also be verified by simulation.

8.2 Strongly Connected Digraphs

In practical wireless settings, not all links may be symmetric due to differing transmit powers (e.g., if the batteries at different nodes have experienced different usage), multipath effects, or interference. To simulate directed networks, we begin with a (undirected) random geometric graph and then add and delete directed edges by random coin flips (while ensuring that the directed graph remains strongly connected). Figure 3 illustrates an example of a strongly connected directed graph. In this example, since the directed graph has fewer edges than the corresponding undirected graph shown in Fig. 1, one would expect that more broadcasts are needed to achieve consensus. The real part of the second smallest eigenvalue of L for the directed graph in Fig. 3 is 0.3930, so the approximately optimal perturbation parameter is $\epsilon^* = 0.1965$.

Fig. 4 illustrates the number of broadcasts with respect to perturbation parameter ϵ required to obtain $\|m(t) - m(t-1)\|_2 \leq 10^{-5}$. The initial values are independent and uniform over $[0, 1]$.

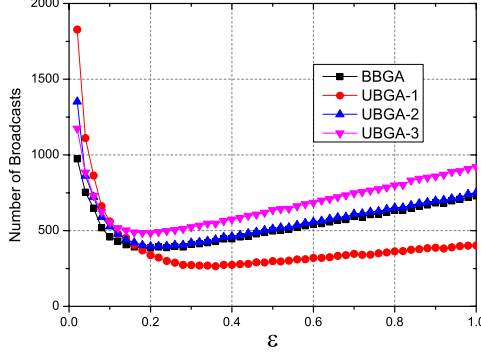


Figure 4: Number of broadcasts to converge with respect to ϵ for simulations on the directed graph shown in Fig. 3.

The optimal perturbation parameter occurs at 0.20, which matches the predicted value well. From this figure, we see that UBGA-1 still gives the best performance

8.3 Scaling Behavior

In the previous two subsections, we illustrate the performance of UBGA and BBGA on particular directed and undirected graphs. In this subsection, we demonstrate the scaling behavior of these two algorithms as the size of the network increases. We compare the performance of three varieties of UBGA and two varieties of BBGA. The three varieties of UBGA are those as defined above, with weights $a_{j,k}$ given in (96), and with $\epsilon = 0.5$. For BBGA, we use the same weights $a_{j,k}$ as given in (70) and set ϵ either to 0.5 (BBGA-0.5) or to ϵ^* (BBGA-opt).

Following [6], we investigate two metrics for error. The UBGA algorithms are guaranteed to converge to the average consensus solution, and we measure the mean squared error after t iterations,

$$r(t) = \frac{1}{n} \|x(t) - \frac{1}{n} \mathbf{1} \mathbf{1}^T x(0)\|_2^2. \quad (97)$$

Since BBGA and other biased broadcast gossip algorithms do not converge to the average consensus, we measure their rate of convergence via the deviation,

$$q(t) = \frac{1}{n} \|x(t) - \frac{1}{n} \mathbf{1} \mathbf{1}^T x(t)\|_2^2, \quad (98)$$

which is guaranteed to go to zero. Note that we do not include the companion variables $y(t)$ in these calculations, since ultimately the aim is to reach consensus only on the states $x(t)$.

As above, we declare that consensus is achieved if $\|m(t) - m(t-1)\|_2 \leq 10^{-5}$. In the sequel, we report the results of numerical simulations on random geometric graphs and random strongly connected digraphs simulated using the procedure described above.

8.3.1 Convergence rate

The number of transmissions required to achieve consensus for undirected graphs and digraphs are shown in Figs. 5 and 6 respectively. Unsurprisingly, BBGA-opt converges significantly faster than BBGA-0.5. However UBGA-1 is still the best one in terms of converge rate performance of scaling behavior. Since UBGA-1 can achieve the average consensus, we prefer UBGA-1 if out-degree information is available; otherwise, BBGA-opt is the winner.

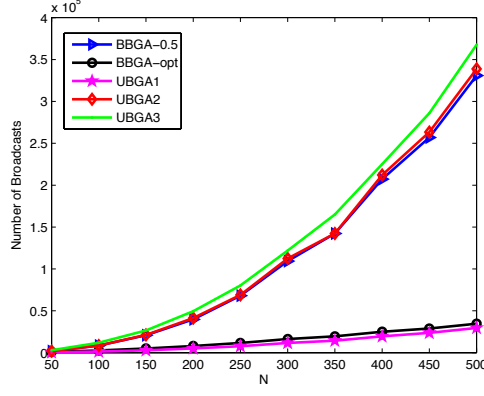


Figure 5: Number of broadcasts to converge with respect to N on undirected graphs.

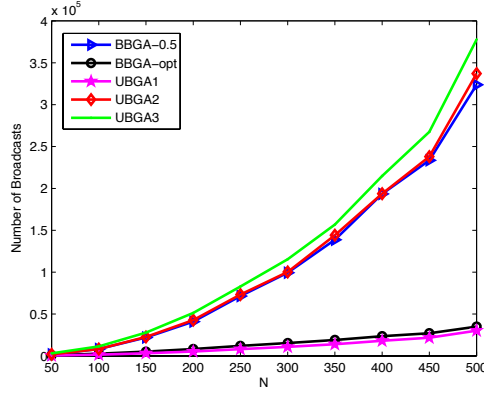


Figure 6: Number of broadcasts to converge with respect to N on digraphs.

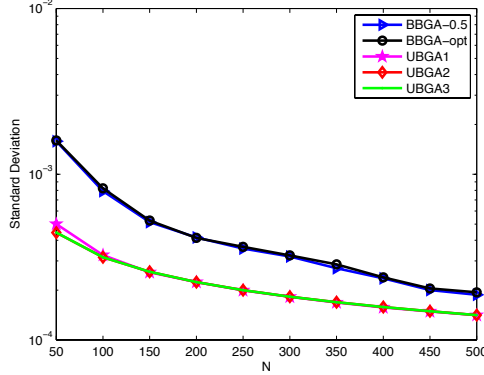


Figure 7: The standard deviation performance with respect to N on undirected graphs.

8.3.2 Deviation

Figures 7 and 8 show the deviation $q(t)$ at the time the algorithm is declared to have converged. For this particular initialization scheme (i.i.d. uniform), we note that the BBGA algorithms are roughly half an order of magnitude worse than the UBGA schemes. We investigate the effects of initialization on deviation further in the next section. For now, we note that both versions of BBGA achieve comparable performance in terms of deviation, and likewise, all three versions of

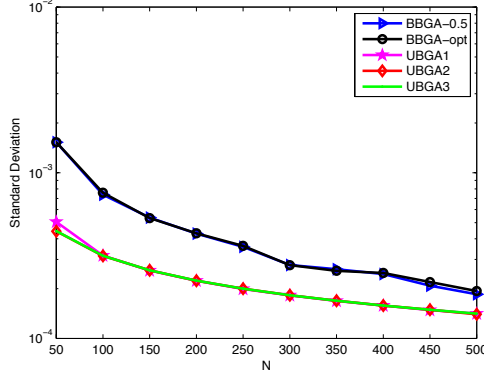


Figure 8: The standard deviation performance with respect to N on digraphs.

UBGA achieve effectively the same deviation at the time they converge.

In summary, from the experiments reported in this section we conclude that UBGA-1 is the most desirable solution if the out-degree information is available (including, when \mathcal{G} is undirected); otherwise BBGA-opt is the next most preferable since it gives the fastest rate of convergence.

9 Performance Analysis

In this section, we compare the broadcast gossip algorithms proposed in this paper with the previous broadcast gossip algorithms of [6] and [7]. In the figures and discussion below, BGA-1 refers to the algorithm in [6], BGA-2 refers to the one in [7], and BBGA and UBGA are the algorithms proposed in this paper. The previous section illustrated that UBGA-1 exhibits many advantages, both in terms of the choice of perturbation parameter and the rate of convergence, compared to the other UBGA algorithms. For this reason, in this section we use UBGA-1 as the representative of the UBGA. For both UBGA and BBGA, we will investigate two settings for the perturbation parameter: $\epsilon = 0.5$ and ϵ^* . Note that ϵ^* is only optimal for BBGA, and it may be suboptimal for UBGA. We also remark that the comparison of BGA-1 and BGA-2 with BBGA-opt and UBGA-opt (i.e., those using ϵ^*) is unfair, since the information used to determine ϵ^* is not made available to either BGA-1 or BGA-2; in particular, neither of those algorithms uses global topology information such as ξ_2 . This is our primary motivation for also considering the performance of UBGA and BBGA with $\epsilon = 0.5$.

All simulations in this section use (undirected) random geometric graph topologies with the same connectivity radius as in the previous section. When \mathcal{G} is directed, BGA-1 is no longer guaranteed to converge to the average consensus in expectation. Unless otherwise noted, each result corresponds to the average over 100 Monte Carlo trials.

Since the initial values effect the performance of various broadcast gossip algorithms, we consider four approaches to initializing the values $x_i(0)$: 1) independent and *uniform* over $[0, 1]$; 2) independent and *Gaussian* with zero mean, unit variance; 3) the *spike* initialization, where one random node has an initial value of 1 and all other nodes have initial values 0; and 4) the *slope* initialization, the initial value at node i is the sum of its x- and y-coordinates in the unit square (note all graphs are drawn from the ensemble of random geometric graphs in the plane).

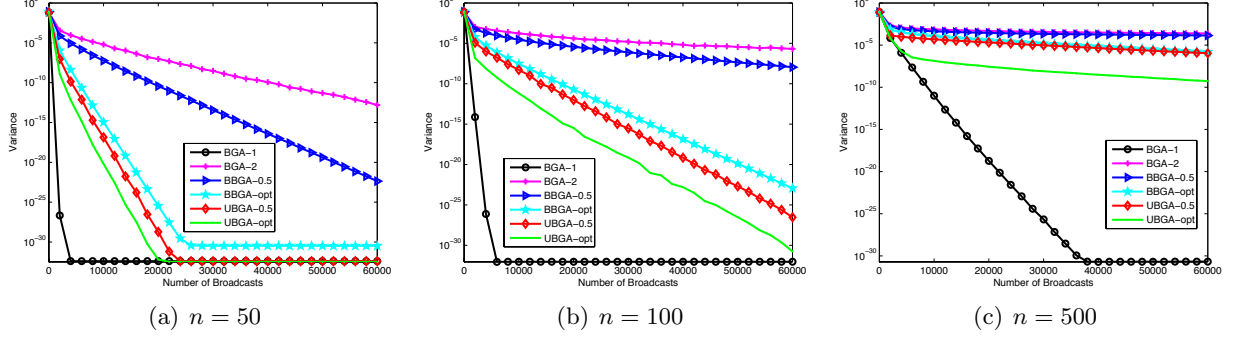


Figure 9: The deviation of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with uniform distribution for initial values.

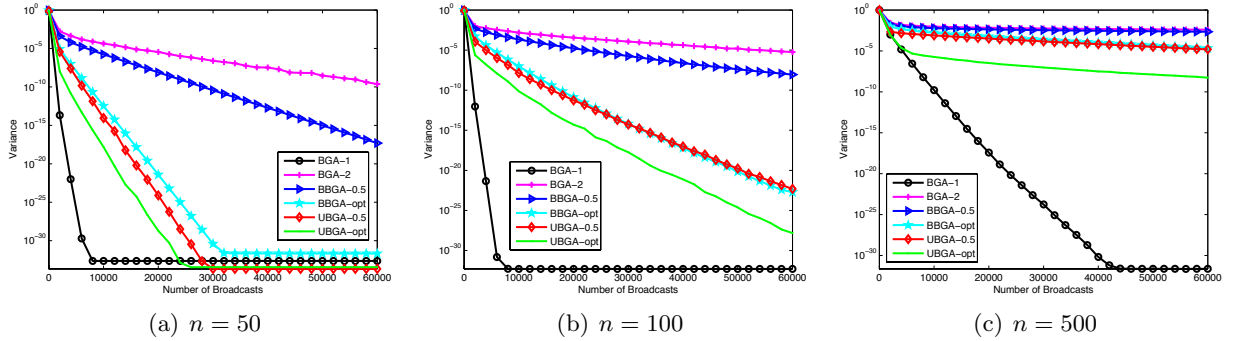


Figure 10: The deviation of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with Gaussian initial values.

9.1 Deviation

Figures 9–12 show the deviation $q(t) = \frac{1}{n} \|x(t) - \frac{1}{n} \mathbf{1} \mathbf{1}^T x(t)\|_2^2$ as a function of time for different initializations. Note that this indicates how quickly the algorithms converge to a consensus, regardless of the value on which consensus is achieved.

It is clear from these figures that BGA-1 converges to its final value faster than the other algorithms. As we will see below, this is because BGA-1 generally achieves a lower accuracy (in terms of mean squared error, $r(t)$) than the other methods.

The algorithms BGA-2, BBGA, and UBGA all maintain companion variables. Among these algorithms we observe that BGA-2 converges slower, in general, than BBGA. Also note that BBGA-opt converges significantly faster than BBGA-0.5 when $n = 50$ or 100 , but the performance of the two is much closer for larger networks. Somewhat surprisingly, the deviation of UBGA-opt and UBGA-0.5 are typically better or comparable to BBGA-opt. This is surprising because it indicates that UBGA is converging faster, despite the fact that it is converging to the average consensus. On the other hand, BGA-1 converges quickly to a consensus which is not on the average, and BGA-2 typically converges to the average consensus but more slowly than UBGA. We conclude that UBGA strikes a desirable balance between converging quickly while achieving consensus on the average.

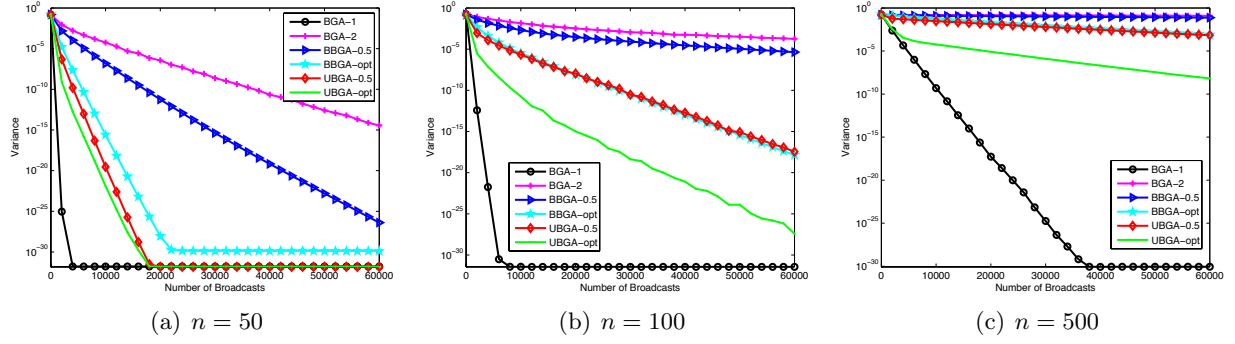


Figure 11: The deviation of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with slope initial values.

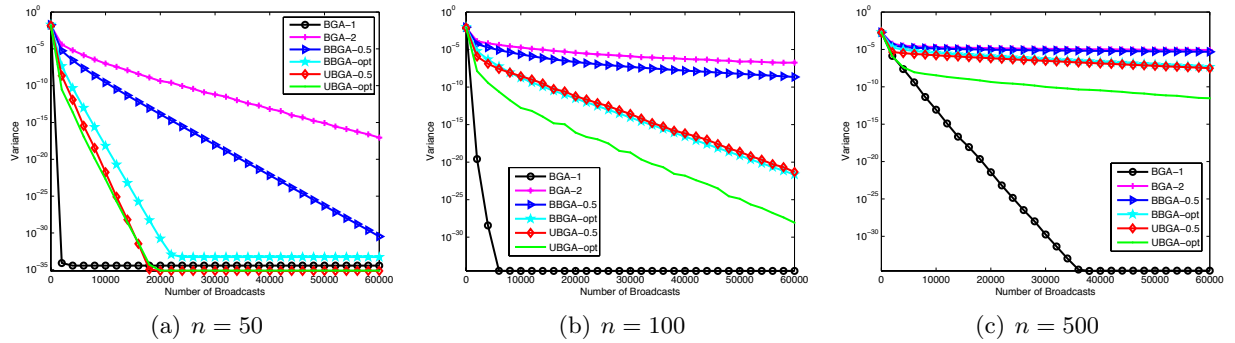


Figure 12: The deviation of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with spike initial values.

9.2 Mean Squared Error

Figs. 13–16 show the mean squared error $r(t) = \frac{1}{n} \|x(t) - \frac{1}{n} \mathbf{1} \mathbf{1}^T x(0)\|_2^2$ as a function of t for all four algorithms and the four different initializations on networks of $n = 50, 100$, and 500 nodes. UBGA generally has the best performance among all algorithms, in the sense that a small deviation is achieved with relatively few broadcasts. BGA-1 has a high deviation; it is well-known that it converges quickly but that it does not converge to the average consensus. When out-degree information is available UBGA is preferable. For networks with $n = 50$ or 100 nodes, using $\epsilon = 0.5$ is close enough to optimal that the performance is extremely good for UBGA-0.5. For larger graphs, the performance of UBGA-opt dominates that of UBGA-0.5. An interesting open problem is to come up with a better practical guideline for setting ϵ as a function of network size and structure, e.g., for random geometric graphs.

It is interesting to note that BBGA has better performance than BGA-2 for a smaller number of broadcasts. Since BGA-2 converges to the average consensus in most examples, but BBGA does not, this indicates that BGA-2 converges slower than BBGA. For larger networks BBGA may be preferable as an alternative which quickly reaches a reasonably accurate solution.

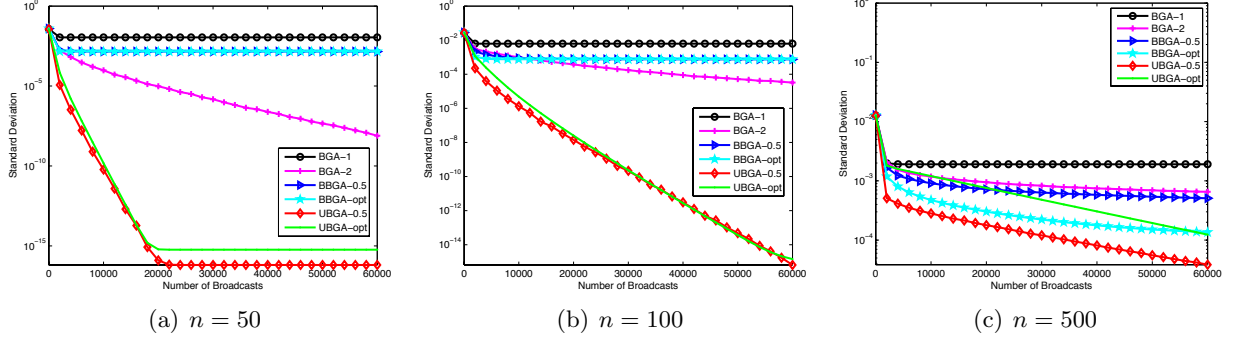


Figure 13: The mean squared error of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with uniform initial values.

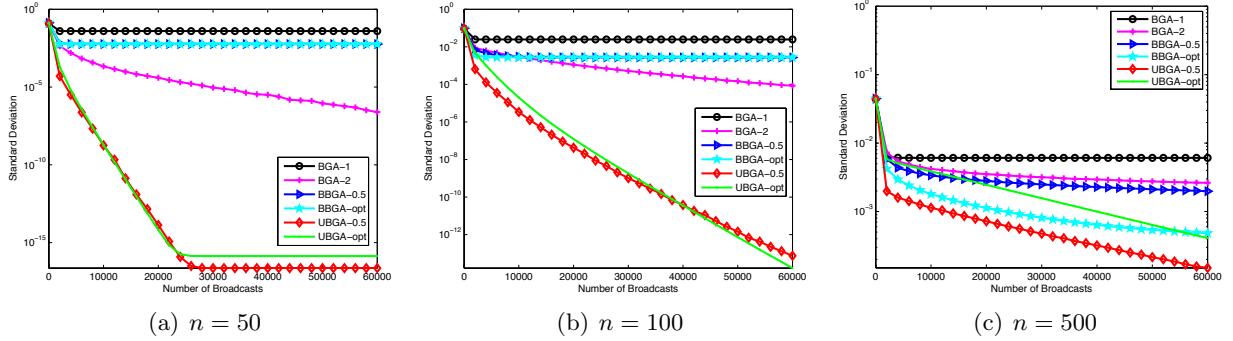


Figure 14: The mean squared error of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with Gaussian initial values.

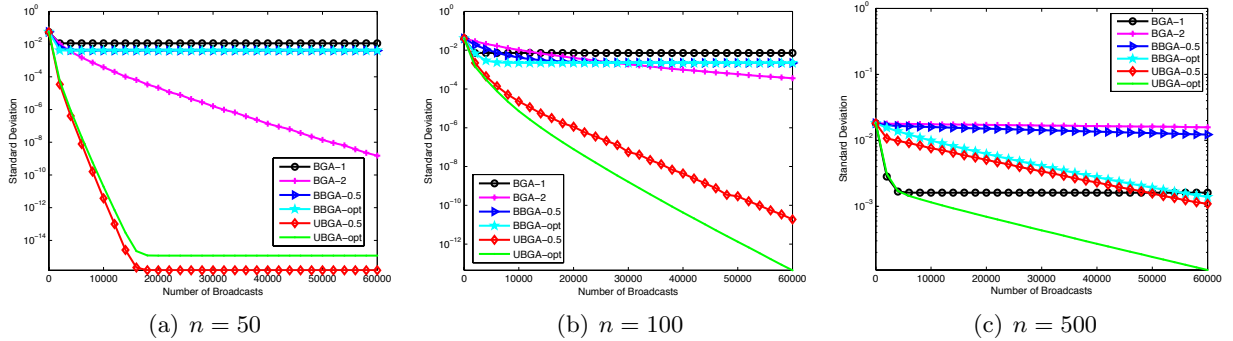


Figure 15: The mean squared error of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with slope initial values.

10 Conclusion and future work

In this paper, we propose a framework for broadcast gossip algorithms and prove that consensus is achieved in both expectation and in the mean squared sense for reasonably chosen coefficients. Then we analyze two particular broadcast gossip algorithms, UBGA and BBGA, where the former preserves the average and the latter is more practical for implementation in non-symmetric broadcast networks. These algorithms have an interpretation from the perspective of matrix per-

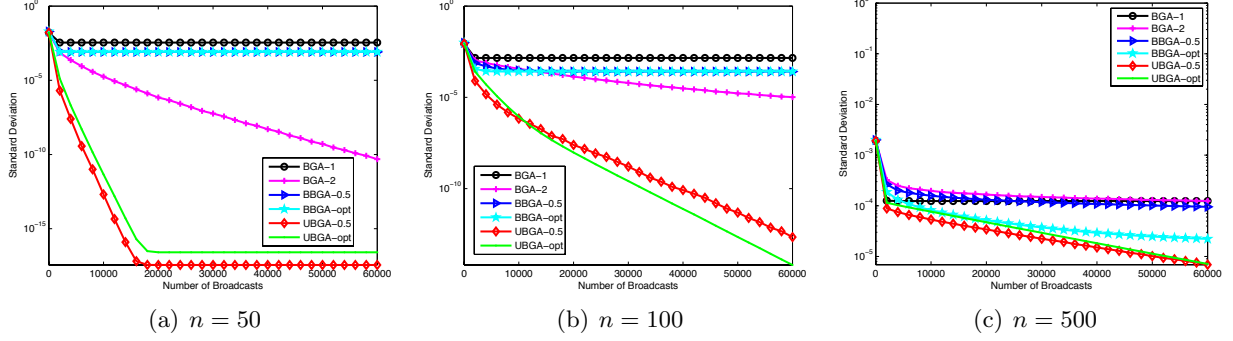


Figure 16: The mean squared error of BGA-1, BGA-2, BBGA and UBGA with respect to the number of broadcasts on undirected random geometric graphs with spike initial values.

turbation, and we derive an upper bound on the perturbation parameter under which convergence is guaranteed. We also study the optimal value of the perturbation parameter and find it is within the range of allowable values. By numerical analysis, the optimal perturbation parameter obtained from BBGA on undirected digraphs is shown to also work well on digraphs. If the out-degree information is available (as is the case in undirected networks), we demonstrate that UBGA outperforms the existing state-of-the-art broadcast gossip algorithms. When out-degree information is not available, BBGA is a promising alternative because it exhibits an excellent tradeoff between the rate of convergence and the limiting mean squared error.

Interesting future work includes studying convergence properties of broadcast gossip algorithms with quantized transmissions. The broadcast gossip algorithms proposed in this paper involve maintaining and transmitting companion variables, in addition to the state variables which are being averaged, and we are interested in understanding how the number of bits allocated to these two different values impacts the rate of convergence and limiting value.

Finally, since wireless media is shared by nodes within communication radius for each other, broadcast packets are likely to undergo collisions and interference, and it would also be interesting to develop a deeper understanding of how broadcast gossip algorithms behave under more realistic channel models (e.g., accounting for capture effects).

References

- [1] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, “Gossip algorithms for distributed signal processing,” *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, November 2010.
- [2] R. Olfati-Saber, J. Fax, and R. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
- [3] M. Mesbahi and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.
- [4] T. Aysal, M. Yildiz, and A. Scaglione, “Broadcast gossip algorithms,” in *Proc. IEEE Information Theory Workshop*, Porto, Portugal, May 2008.
- [5] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, “Broadcast gossip algorithms: Design and analysis for consensus,” in *Proc. IEEE Conf. Decision and Control*, Cancun, Mexico, Dec. 2008.

- [6] —, “Broadcast gossip algorithms for consensus,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, July 2009.
- [7] M. Franceschelli, A. Giua, and C. Seatzu, “Consensus on the average on arbitrary strongly connected digraphs based on broadcast gossip algorithms,” in *Proc. 1st IFAC Workshop on Estimation and Control of Networked Systems*, 2009, pp. 66–71.
- [8] J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Trans. Automatic Control*, vol. 31, no. 8, pp. 803–812, Sep. 1986.
- [9] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [10] F. Fagnani and P. Frasca, “Broadcast gossip averaging: Interference and unbiasedness in large Abelian Cayley networks,” *IEEE J. Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 866–875, August 2011.
- [11] A. Nedić, “Asynchronous broadcast-based convex optimization over a network,” *IEEE Trans. Automatic Control*, vol. 56, no. 6, pp. 1337–1351, June 2011.
- [12] M. Franceschelli, A. Giua, and C. Seatzu, “Distributed averaging in sensor networks based on broadcast gossip algorithms,” *IEEE Sensors Journal*, vol. 11, no. 3, pp. 808–817, March 2011.
- [13] K. Cai and H. Ishii, “Average consensus on general digraphs,” in *Proc. IEEE Conf. Decision and Control*, Orlando, FL, 2011, pp. 1956–1961.
- [14] —, “Average consensus on general strongly connected digraphs,” March 2012, available at <http://arxiv.org/abs/1203.2563>.
- [15] C. Meyer, *Matrix analysis and applied linear algebra*. SIAM, 2001.
- [16] A. Seyranian and A. Mailybaev, *Multiparameter stability theory with mechanical applications*. World Scientific, 2004.
- [17] E. Seneta, *Non-negative Matrices and Markov Chains*. Springer, 1981.
- [18] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, September 2004.
- [19] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.
- [20] R. Bhatia, *Matrix Analysis*. Springer-Verlag, 1996.
- [21] G. Stewart and J. Sun, *Matrix Perturbation Theory*. Academic Press, 1990.
- [22] A. Berman and X. Zhang, “Lower bounds for the eigenvalues of Laplacian matrices,” *Linear Algebra and its Applications*, vol. 316, no. 1–3, pp. 13–20, September 2000.
- [23] P. Gupta and P. R. Kumar, “Critical power for asymptotic connectivity in wireless networks,” in *Stochastic Analysis, Control, Optimization, and Applications*, Boston, 1998, pp. 1106–1110.
- [24] M. Penrose, *Random Geometric Graphs*. Oxford University Press, 2003.